Structural insight into the biosynthesis of molybdenum cofactor and heme d_1

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktor der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
Dissertation

von Sayantan Saha

aus Siliguri, Indien

1. Referent: Prof. Dr. Dirk Heinz

2. Referent: Prof. Dr. Ralf Mendel

eingereicht am: 27.05.2013

mündliche Prüfung (Disputation) am: 12.08.2013

Druckjahr 2014

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikation

Storbeck, S., Saha, S., Krausze, J., Klink, B.U., Heinz, D.W. & Layer, G. (2011) Crystal structure of the heme d_1 biosynthesis enzyme NirE in complex with its substrate reveals new insights into the catalytic mechanism of S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferases *J. Biol. Chem.* **286,** 26754-26767.

Tagungsbeiträge

Saha, S., Bittner, F., Schmidt, N., Kruse, T., Mendel, R.R. & Heinz, D. W. : Structure function studies of molybdenum enzymes. (Poster) International PhD Symposium, Braunschweig, Germany, (December 2009).

Saha, S., Bittner, F., Schmidt, N., Kruse, T., Mendel, R.R. & Heinz, D. W. : Structural analysis of insertion of sulfurated molybdenum cofactor into target enzymes. (Poster) International School of Crystallography, Erice, Italy, June 2010.

Saha, S., Storbeck, S., Krusze, J., Klink, B. U., Heinz, D. W. & Layer, G. : Crystal structure of NirE in complex with its substrate reveals new insights into the catalytic mechanism of SUMT enzymes. (Talk) DGK Meeting, Salzburg, Austria, September 2011.

Saha, S., Bittner, F., Schmidt, N., Kruse, T., Mendel, R.R. & Heinz, D. W. : Atomic structures of proteins related to the insertion of molybdenum cofactor and heme. (Talk) International Workshop "Prosthetic groups transport and insertion", Warberg, Germany, September 2011.

Saha, S., Kruse, T., Mendel, R.R. & Heinz, D. W. : Crystal structure of Cnx1 E domain gives insight into the final step of Moco Biosynthesis. (Talk) 14th Heart of Europe Bio-Crystallography Meeting, Zagan, Poland, October 2011.

Saha S., Storbeck, S., Krausze, J., Klink, B. U. & Heinz, D. W., Layer, G. : Crystal mechanism of SUMT enzymes. (Poster) International PhD Symposium, Braunschweig, Germany, December 2012.

Table of Contents

Summary

Almost all life processes ranging from respiration to metabolism are controlled and coordinated by enzymatic reactions. A large repertoire of these enzymes requires additional components known as cofactors to achieve optimal activity. Organic cofactors or prosthetic groups like heme, siroheme, NAD^+ etc. directly participate in the enzymatic reactions, thereby bringing about challenging biocatalytic transformations. While the mechanisms of reactions involving cofactors are being studied for decades, their biosynthesis, transport and insertion into target enzymes remain largely unknown. This study focuses on the synthesis of two such prosthetic groups, the molybdenum cofactor (Moco) and heme d_1.

Moco and heme d_1 are synthesized via evolutionarily conserved multistep pathway involving various enzymatic reactions. This work deals with the structural investigation of two enzymes, Cnx1E and NirE, which are implicated in the synthesis of Moco and heme d_1, respectively. Cnx1E is the N-terminal domain of the protein Cnx1 and catalyzes the final step of Moco biosynthesis, whereas NirE is involved in the initial transformation of uroporphyrinogen III (uro'gen III) to precorrin-2, which after several further modification leads to the formation of heme d_1. As a part of this investigation, both the proteins were recombinantly expressed in *E. coli*, purified by affinity chromatography and crystallized to obtain single diffracting crystals. While Cnx1E could be crystallized only in its apo form, NirE crystallized in apo as well as substrate bound forms. The crystals were subjected to X-ray crystallographic analysis to solve the molecular structure of the proteins.

The structural analysis coupled with functional characterization of NirE and its variants concluded in the elucidation of its detailed enzymatic mechanism where two arginine residues were identified as catalytically active base and initiator of the methylation of uro'gen III, thereby converting the substrate to precorrin-2. The structure of Cnx1E revealed its four well-separated domain organization. An interesting discovery made from this structure was the high degree of structural similarity of sub-domain 3 of Cnx1E to the C-terminal domain of Cnx1 (Cnx1G) leading to the hypothesis that sub-domain 3 harbors the actual catalytic site of the enzyme. This hypothesis was further explored by mutational analysis of two previously identified, functionally relevant residues. In order to reach a definite conclusion regarding the enzymatic mechanism of Cnx1E, it is imperative to elucidate the structure of enzyme-substrate complex. This needs further investigation.

Zusammenfassung

Fast alle Prozesse des Lebens wie Metabolismus und Zellatmung werden durch enzymatische Reaktionen koordiniert und kontrolliert. Viele Enzyme benötigen zusätzliche Moleküle, Kofaktoren genannt, für die optimale Aktivität. Organische Kofaktoren oder prosthetische Gruppen wie Häm, Sirohäm, NAD^+ partizipieren direkt in enzymatischen Reaktionen, um komplizierte biokatalytische Umwandlungen zu ermöglichen. Während die Reaktionsmechanismen, in die Kofaktoren involviert sind, oft seit Jahrzehnten erforscht werden, ist über deren Biosynthese, Transport und Einbau in ihr Enzym meist sehr wenig bekannt. Diese Arbeit befasst sich mit der Synthese zweier solcher prosthetischer Gruppen, dem Molybdänkofaktor (Moco) und Häm d_1.

Moco und Häm d_1 werden in mehreren Schritten hergestellt über einen evolutionär konservierten Syntheseweg, der verschiedene enzymatische Reaktionen beinhaltet. In dieser Arbeit wurden die Strukturen von zwei Enzymen, Cnx1E and NirE, aufgeklärt, die in der Synthese von Moco and heme d_1 involviert sind. Cnx1E ist die N-terminale Domäne des Proteins Cnx1 und katalysiert den finalen Schritt in der Moco Biosynthese, während NirE an der initialen Umwandlung von Uroporphyrinogen III (Uro'gen III) zu Precorrin-2 beteiligt ist, welches durch weitere Reaktionsschritte zu Häm d_1 umgwandelt wird. In dieser Arbeit wurden beide Proteine rekombinant in *E. coli* exprimiert, über Affinitätschromatographie gereinigt und kristallisiert. Während Cnx1E nur in der apo-Form kristallisiert werden konnte, wurde NirE sowohl ohne als auch mit gebundenem Kofaktor kristallisiert werden. Die molekulare Struktur der beiden Proteine wurde mittels Röntgenkristallographie aufgeklärt.

Die Strukturaufklärung zusammen mit der funktionellen Charakterisierung von NirE und Varianten davon erlaubte die Aufklärung des enzymatischen Mechanismus. Dabei agieren zwei Argininreste als katalytisch aktive Base und als Initiator der Methylierung von Uro'gen III, wodurch das Substrat zu Precorrin-2 umgewandelt wird. Die Struktur von Cnx1E zeigte, dass das Protein in vier separate Domänen aufgeteilt ist. Interessanterweise ist die Struktur der Domäne 3 sehr ähnlich zu der C-terminalen Domäne von Cnx1 (Cnx1G), was zu der Hypothese führt, dass die Domäne 3 für die tatsächliche katalytische Aktivität zuständig ist. Diese Hypothese wurde weiter untersucht durch die Analyse von zwei Mutanten von zwei bekannten, funktionell

relevanten Aminosäuren. Um jedoch eine definitive Aussage über den Reaktionsmechanismus von Cnx1E treffen zu können, wäre es notwendig die Struktur von Cnx1E im Komplex mit seinem Substrat aufzuklären.

Abbreviations

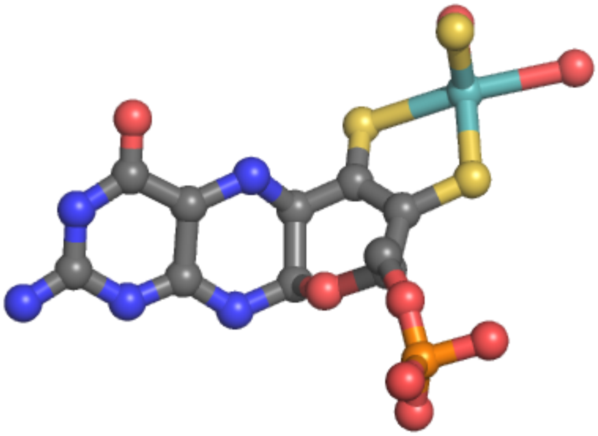
3D	3 Dimensional
Aba3	(+)-abscisic acid 8'-hydroxylase
ALA	5-aminolevulinic acid
AMP	Adenosine monophosphate
AMP	Adenosine mono phosphate
AP	Alkaline phosphate
APS	Ammonium persulfate
Cnx	Cofactor for nitrate reductase and xanthine dehydrogenase
CobA	Cobalt-precorrin 5A hydrolase
cPMP	cyclic pyranopterin
CysG	5-methyltetrahydrofolate--homocysteine S-methyltransferase
ddH_2O	Distilled water
DMSOR	Dimethyl sulfoxide reductase
E. Coli	Escherichia coli
FAD	Flavin adenine dinucleotide
FeMoco	Iron molybdenum cofactor
GTP	Guanosine-5'-triphosphate
H_2SO_4	Sulfuric acid
HCl	Hydrochloric acid
Hepes	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HPLC	High pressure liquid chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
KI	Kotassium iodide
LB	Luria-Bertani broth
MCP	Moco carrier protein
Met1	Methionine S-methyltransferase
$MgCl_2$	Magnesium chloride
Moco	Molybdenum cofactor
MPT	Molydopterin
MPT	Molydopterin
NaCl	Sodium chloride
NAD	Nicotinamide adenine dinucleotide
NirE	Nitrite reductase protein encoded by *nirE* gene
PDB	Protein data bank
PEG	Polyethylene glycol
RLD	Rhodonase like domain
SAH	S-adenosyl-homocysteine
SAM	S-adenosyl-L-methionine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SirA	Sirohydrochlorin ferrochelatase
SUMT	SAM dependent uroporphyrinogen III methyl transferase
TEMED	Tetramethylethylenediamine
TLS	Translation Libration Screw-motion
Tris	Trisaminomethane

ABBREVIATIONS

E. coli	Escherichia coli
MoO_4^{2-}	Molybdate
[Fe-S]	Iron sulfur cluster
ATP	Adenosine triphosphate
Mg	Magnesium
Ni	Nickel
Co	Cobalt
NO	Nitric oxide
Cu	Copper
Fe	Iron

1. Introduction

1A. Role of Cnx1E in the synthesis of molybdenum cofactor



1A.1. Molybdenum Enzymes

Molybdenum is one of the most abundant transition elements on Earth. Although it makes up only a small fraction of the earth's crust, it is readily available in sea water and as molybdate (MoO_4^{2-}) (Hille, 2002). Upon cellular uptake, MoO_4^{2-} is incorporated into metal cofactors by a conserved biosynthetic pathway (Dos Santos *et al*., 2004, Schwarz, 2005). These cofactors form an essential component of the active site of molybdenum enzymes that play a cardinal role in the biogeochemical redox cycles of carbon, nitrogen and sulfur on earth, as well as in the metabolism of every individual organism. The versatile redox chemistry of molybdenum is extensively exploited by these enzymes to catalyze redox reactions (Hille, 2002). This is essentially achieved by the three oxidation states of molybdenum (+4, +5 and +6), thereby enabling the enzymes to catalyze two-electron reduction/oxidation reactions. With the exception of *Saccharomyces cerevisiae*, molybdenum enzymes are found in all organisms (Zhang *et al*., 2008). Up till now, more than 50 molybdenum enzymes have been discovered.

Based on the catalytic functions and on the type of cofactor, these enzymes are broadly divided into two categories. The first category consists of the iron-molybdenum cofactor (FeMoco) containing bacterial nitrogenases and the second group is composed of enzymes harboring pterin-based cofactors (Moco). Both these types of molybdenum cofactors share similar metal ion coordination where molybdenum is coordinated by three sulfur and two oxygen atoms. The first type, FeMoco, is an iron-sulfur-cluster based iron-molybdenum cofactor where one [4Fe-3S] and one [Mo-3Fe-3S] are linked by a partial cubane bridge of three sulfides and one μ_6 central atom, X (which could be carbon, oxygen or nitrogen) (Figure 1A) (Einsle*et al.*, 2004). The other type of molybdenum cofactor, known as Moco, is a pterin based cofactor, which contains a C-6 substituted pyrano ring, a terminal phosphate and a unique dithiolate group that participates in the coordination of molybdenum (Rajagopalan & Johnson, 1992). The metal can be attached to one or two pterin moieties with additional terminal oxygen and sulfur ligands. The latter category of enzymes is further classified into three families: dimethyl sulfoxide reductase (DMSOR), sulfite oxidase and xanthine oxidase. Each of these enzyme families has a distinct active-site structure (Hille, 2002). The chemical and three-dimensional structures of different molybdenum cofactors are shown in Figure 1. This work focuses on the biosynthesis of Moco belonging to the xanthine oxidase family of plant enzymes.

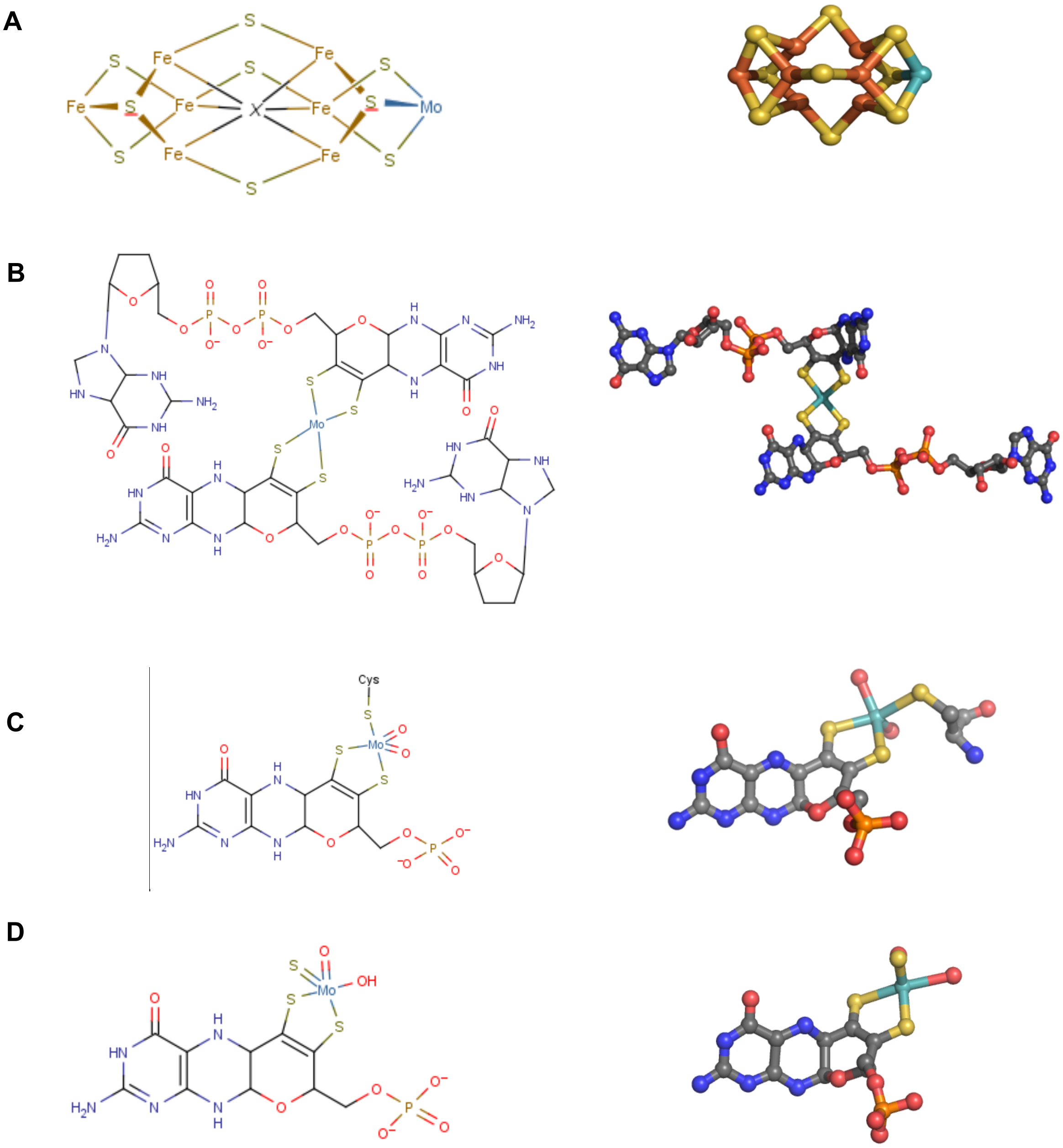


Figure 1: Molybdenum cofactors. The structures of four different types of molybdenum cofactors, which are the basis of classification of molybdenum enzymes, are shown. Schematics are shown on the left-hand side, full 3D structure are represented by balls-and-sticks on the right-hand site. (A) FeMoco of molybdenum nitrogenase from *Azobacter vinelandii* (PDB ID–1N2C), (B) Moco of DMSOR from *Rhodobacter sphaeroides* (1EUI) (C) Moco of chicken sulphite oxidase (3R18), (D) Moco of bovine xanthine oxidase (1FIQ).

Xanthine oxidase family of enzymes - Enzymes belonging to the xanthine oxidase family are also known as molybdenum hydroxylases due to their catalytic function in hydroxylation reaction. They are known to catalyze the hydroxylation of the carbon centers in aldehydes and aromatic heterocycles (Hille, 2005). Members of this family are represented by xanthine oxidase, xanthine dehydrogenase and aldehyde oxidase. These enzymes play a key role in purine degradation (Schwarz *et al*., 2009). In plants, they are also believed to be involved in pathogen defense (Montalbini, 1992) and ageing (Pastori and Del Rio, 1997). The crystal structures of several members of this enzyme family have been elucidated including bovine xanthine oxidase (Enroth *et al.,* 2000) and human xanthine oxidase (Yamaguchi *et al.,* 2007). These structures revealed that eukaryotic xanthine oxidase and xanthine dehydrogenase assemble as homodimers having three distinct domains in each monomer. Apart from Moco, each of the monomers also harbors two iron-sulfur centers and one FAD cofactor for electron transfer. The Moco binding domain is involved in and is essential for the dimerization and hence constitution of the enzyme. Figure 2 shows the structure of the homodimeric bovine xanthine oxidase along with the three different cofactors.

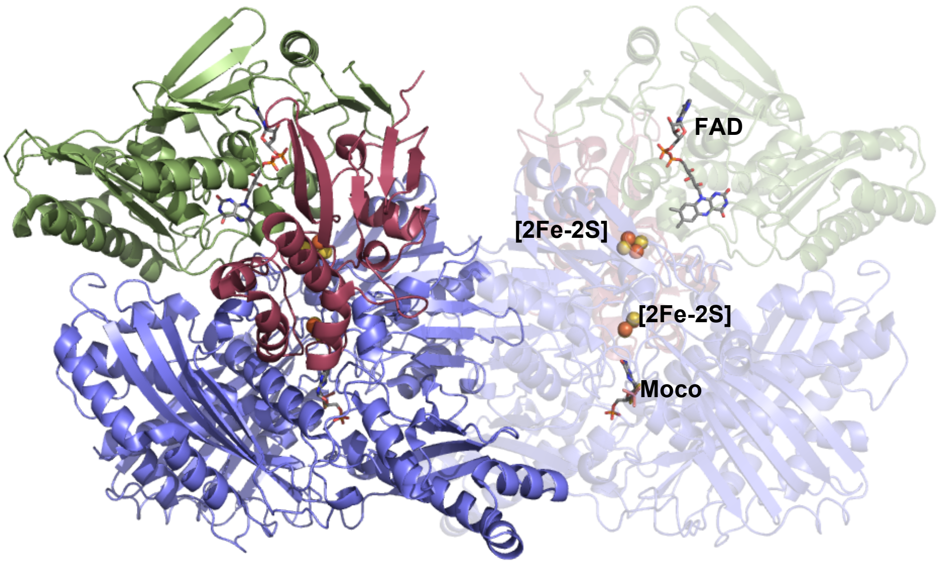


Figure 2: Crystal structure of xanthine oxidase from bovine milk. The homodimer of bovine xanthine oxidase is shown in cartoon representation (PDB ID-1FIQ). One of the monomers is shown with higher transparency to provide a clear view of the cofactors that are buried deep within the molecule. FAD and Moco are shown as sticks whereas the iron sulfur clusters are shown as spheres. The three different domains of each subunit are differentiated by colors. The FAD binding domain is shown in green, the domain harboring the iron sulfur clusters is shown in red and the Moco binding domain, which is also the dimerizing domain, is shown in blue (Enroth *et al*., 2000).

1A.2. Moco biosynthesis

Moco biosynthesis takes place through a highly conserved pathway comprising of four major steps (Schwarz, 2005). It starts with GTP and involves the formation of the intermediates precursor Z, molybdopterin (MPT), adenylated MPT (MPT-AMP) and finally Moco (Schwarz, 2005). Figure 3 depicts the four major steps of the Moco biosynthetic pathway in plants, along with the key enzymes involved in the numerous catalytic reactions. These enzymes share a common **Cnx** nomenclature that is derived from their overall function in the synthesis of **c**ofactor for **n**itrate reductase and **x**anthine dehydrogenase. Moco is a very labile and oxygen sensitive molecule. Once synthesized, it cannot exist in a free state and has to be coupled to another protein (Rajagopalan & Johnson, 1992). Thus, it is either transferred to the molybdenum enzymes or is complexed to a carrier protein (MCP), which stores the cofactor until further use. Another possibility at this stage is to transfer Moco to the Moco sulfurase protein (Aba3), which further sulfurates it before it can be used as a cofactor by enzymes of the xanthine oxidase family (Bittner *et al.,* 2001, Mendel and Bittner, 2006). The individual steps of Moco biosynthesis in plants are discussed below in detail.

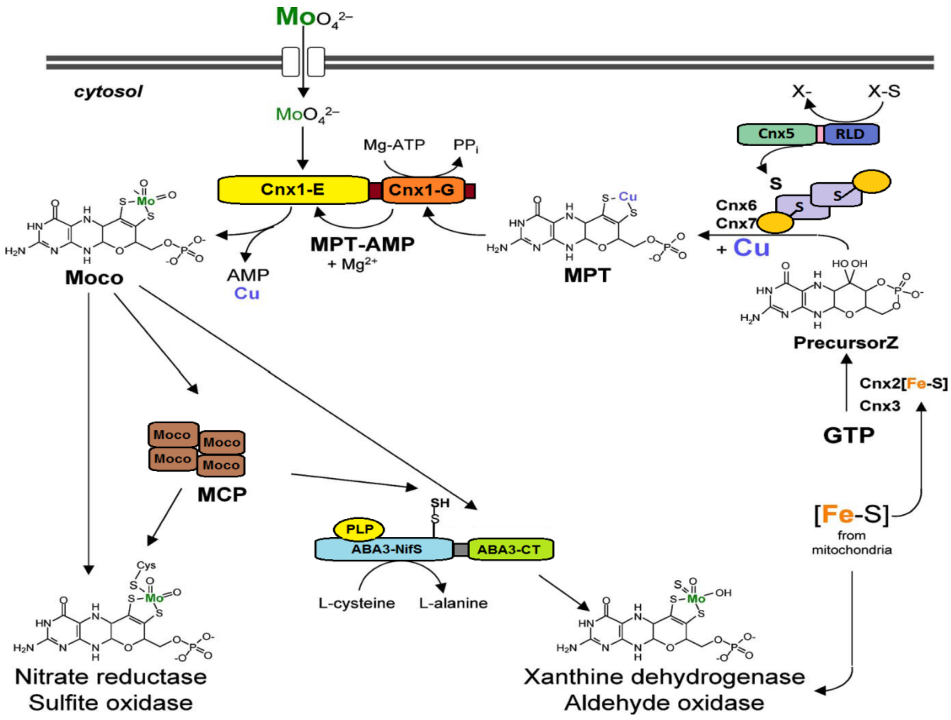


Figure 3: Moco Biosynthesis pathway, distribution and maturation in plant cell. The steps of Moco biosynthesis pathway starting from GTP to Moco, along with the intermediates and the enzymes catalyzing the individual steps are depicted. (Figure modified from Mendel and Bittner, 2006)

Conversion of GTP to cyclic pyranopterin (cPMP) - Moco biosynthesis starts with the conversion of GTP to cyclic pyranopterin (cPMP), which was initially named precursor Z (Wuebbens and Rajagopalan, 1993). In plants, this step is catalyzed by two enzymes, Cnx2 and Cnx3 (Figure 4). Cnx2 is a S-adenosyl-L-methione (SAM) enzyme and harbors two [4Fe-4S] clusters, one at the N-terminal end and the other at the C-terminal end. The [4Fe-4S] at the N-terminal end is involved in the generation of radical SAM and the C-terminal [4Fe-4S] cluster is involved in substrate binding (Hanzelmann and Schindelin, 2006). The second enzyme (Cnx3) is a hexameric protein and plays a role in the release of the pyrophosphate. The detailed reaction mechanism of this step is a subject of current research and a conclusive understanding regarding the enzymatic catalysis is yet to be reached. However, it has been proven that cPMP contains all the carbon atoms of the GTP-derived ribose and purine (Hanzelmann and Schindelin, 2006, Wuebbens and Rajagopalan, 1995). The cPMP molecule is a completely reduced pyranopterin with a terminal cyclic phosphate and a geminal diol.

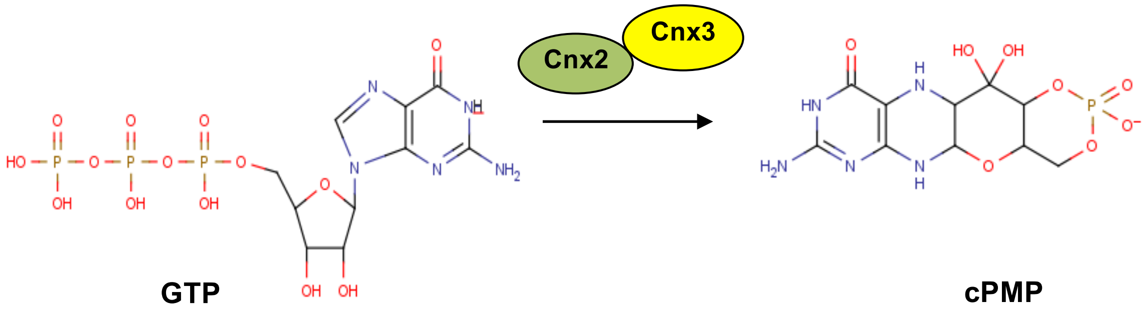


Figure 4: Conversion of GTP to cyclic pyranopterin (cPMP). Schematic representation of the conversion of GTP to cPMP, catalyzed by the enzymes Cnx2 and Cnx3 in plants.

Synthesis of molybdopterin dithiolate (MPT) - The second step in Moco biosynthesis is the conversion of cPMP to molybdopterin dithiolate (MPT) catalyzed by the enzyme MPT synthase (Figure 5). MPT synthase is a heterotetrameric assembly of two small and two large subunits (Schwarz *et al.*, 2009). In plants, this tetrameric assembly has been suggested to be composed of Cnx6 (small subunit) and Cnx7 (large subunit) (Schwarz, 2005). For the formation of MPT dithiolate, two sulfur atoms need to be incorporated into cPMP. These two sulfur atoms come from the two small subunits where it is bound at the C-terminal double glycine motif (Gutzke *et al.*, 2001). The original source of the sulfur atoms is not yet known. This motif is buried deep inside the large subunit to form the active center of the enzyme (Rudolph *et al.*, 2001). Owing to the presence of only one sulfur atom per molecule, this reaction has been proposed to proceed through a two-step mechanism via a monosulfurated intermediate (Wuebbens *et al.*, 2003). An unexpected discovery regarding the MPT molecule synthesized at this step, was the existence of a bound copper to the dithiolate sulfurs (Kuper *et al.*, 2004).

Once the sulfur atoms have been transferred to cPMP, the desulfo-Cnx7 is catalytically regenerated by MPT synthase sulfurase (Figure 5). This is brought about by the adenylation of the C-terminal glycine residue and is followed by sulfuration by a cysteine desulfurase. In plants, these two steps of adenylation and sulfuration are performed by the enzyme Cnx5 (Matthies *et al.*, 2004, Matthies *et al.*, 2005). Cnx5 is a two domain protein where the N-terminal domain is responsible for the adenylation of MPT sulfurase, followed by sulfuration by the C-terminal rhodanese-like protein (RLD) (Matthies *et al.*, 2005).

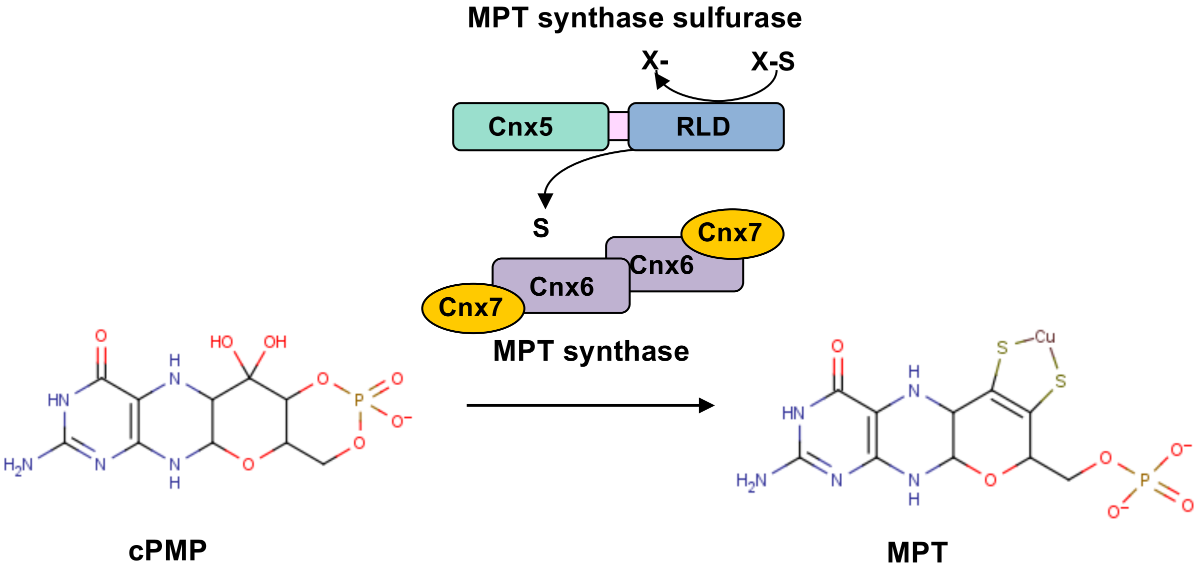
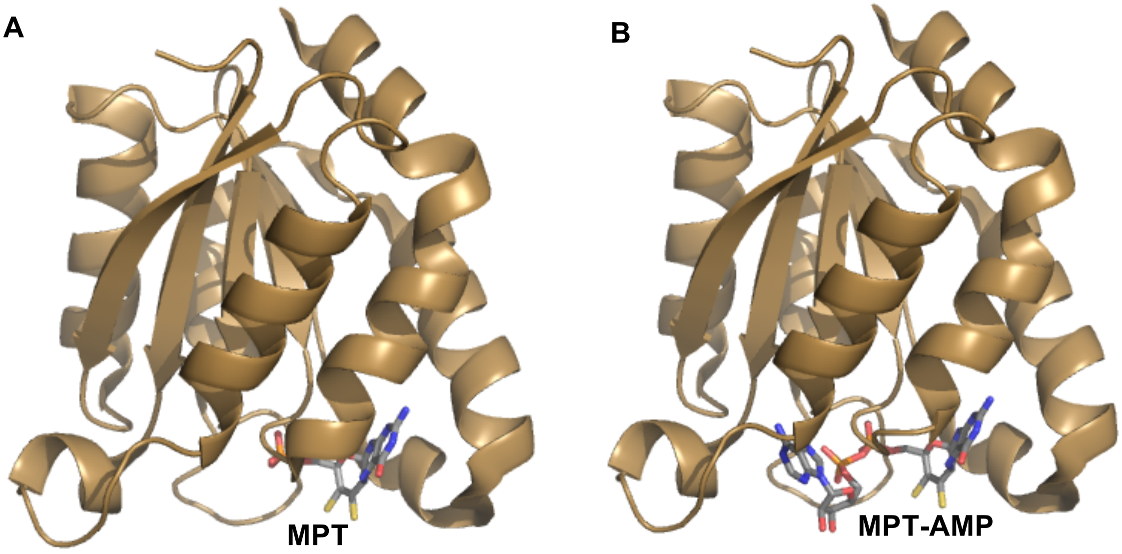


Figure 5: Conversion of cPMP to MPT. The cPMP is converted to MPT by the enzyme MPT synthase which is a heterotetrameric complex of two large subunits Cnx6 and two small subunits Cnx7. Each molecule of Cnx7 has one sulfur atom bound. Following each sulfuration by MPT synthase, it is regenared by MPT synthase sulfurase.

Adenylation of MPT - The synthesis of MPT completes the skeleton of the cofactor. The next step involves the insertion of the molybdenum metal into MPT to form Moco. Hence, this step is the crosslink between the MoO_4^{2-} uptake system and the MPT synthesis pathway. Molybdenum is incorporated into MPT by a two step reaction. In bacteria, two enzymes, MogA and MoeA are involved in the two steps of this reaction (Schwarz, 2005). However, upon evolution to eukaryotes, these two enzymes have been fused to form one single protein molecule with two domains. In plants, this enzyme is known as Cnx1 (Schwarz *et al*., 1997, Schwarz *et al*., 2001). Cnx1 is composed of a C-terminal G-domain (Cnx1G) and an N-terminal E-domain (Cnx1E). The names of the domains Cnx1**G** and Cnx1**E** are derived from the bacterial homologues Mo**g**A and Mo**e**A, respectively. The structures of wildtype Cnx1G and Cnx1G variants in complex with its substrate and product have already been solved (Kuper *et al*., 2004, Kuper *et al.,* 2006).

The structure of wildtype Cnx1G in complex with its substrate MPT (PDB ID-1UUX), revealed the existence of the copper molecule bound to the dithiolate sulfurs. The exact role of this copper is yet to be ascertained. It has been speculated that the copper atom is involved in protecting the MPT dithiolate from oxidation. Additionally, it could also be possible that the copper atom presents an appropriate exit group for the insertion of molybdenum (Mendel and Bittner, 2006).

The physiological function of the Cnx1 enzyme was elucidated from the X-ray crystallographic analysis of Cnx1G variants (Kuper *et al*., 2006). An important discovery made from the structure of Cnx1G variant S583A (PDB ID-1UUY), was the existence of the novel intermediate MPT-AMP bound at the catalytic site. The adenosine molecule was found to be linked to the C-4 carbon of the MPT moiety via a phosphate bond. Hence, the role of Cnx1G was established as an adenylation enzyme that activates MPT by adenylation to form MPT-AMP in Mg^{2+} and ATP dependent manner. Figure 6A shows the crystal structure of Cnx1G in complex with its substrate, MPT, whereas Figure 6B shows the structure of Cnx1G S583A variant in complex with its product, MPT-AMP. The chemical reaction catalyzed by Cnx1G is depicted in Figure 6C.



Enzyme substrate complex

Enzyme product complex

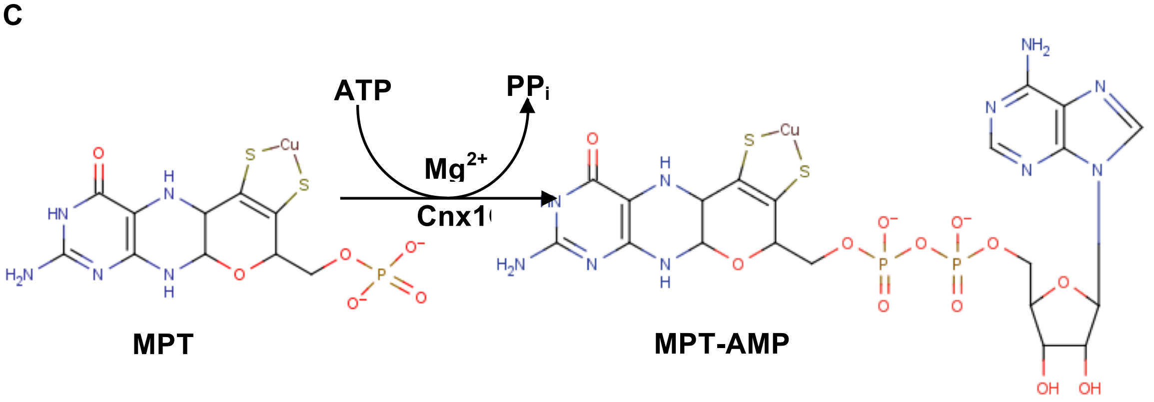


Figure 6: Crystal structures of and the reaction catalyzed by Cnx1G from *Arabidopsis thaliana*. Structure of Cnx1G in complex with its substrate MPT (PDB ID-1UUX)(A) and the product MPT-AMP (1UUY) (B). In both the figures, Cnx1G is shown in cartoon representation whereas the substrate MPT and the product, MPT-AMP bound at the active site are represented by sticks (Kuper *et al.*, 2004). (C) Adenylation of MPT to MPT-AMP catalyzed by Cnx1G.

Insertion of molybdenum and crosstalk with copper metabolism - In the fourth and final step of the Moco biosynthesis pathway, MPT-AMP is converted to mature Moco. MPT-AMP formed in the previous step by the C-terminal Cnx1G is transferred to the N-terminal Cnx1E, thus establishing a product-substrate channel (Figure 7). Cnx1E hydrolyses MPT-AMP, releases the copper and subsequently inserts a molybdenum atom in the form of MoO_4^{2-}, thereby leading to the formation of mature Moco (Schwarz *et al.*, 2009). The structural analysis of Cnx1E is one of main objectives of this work.

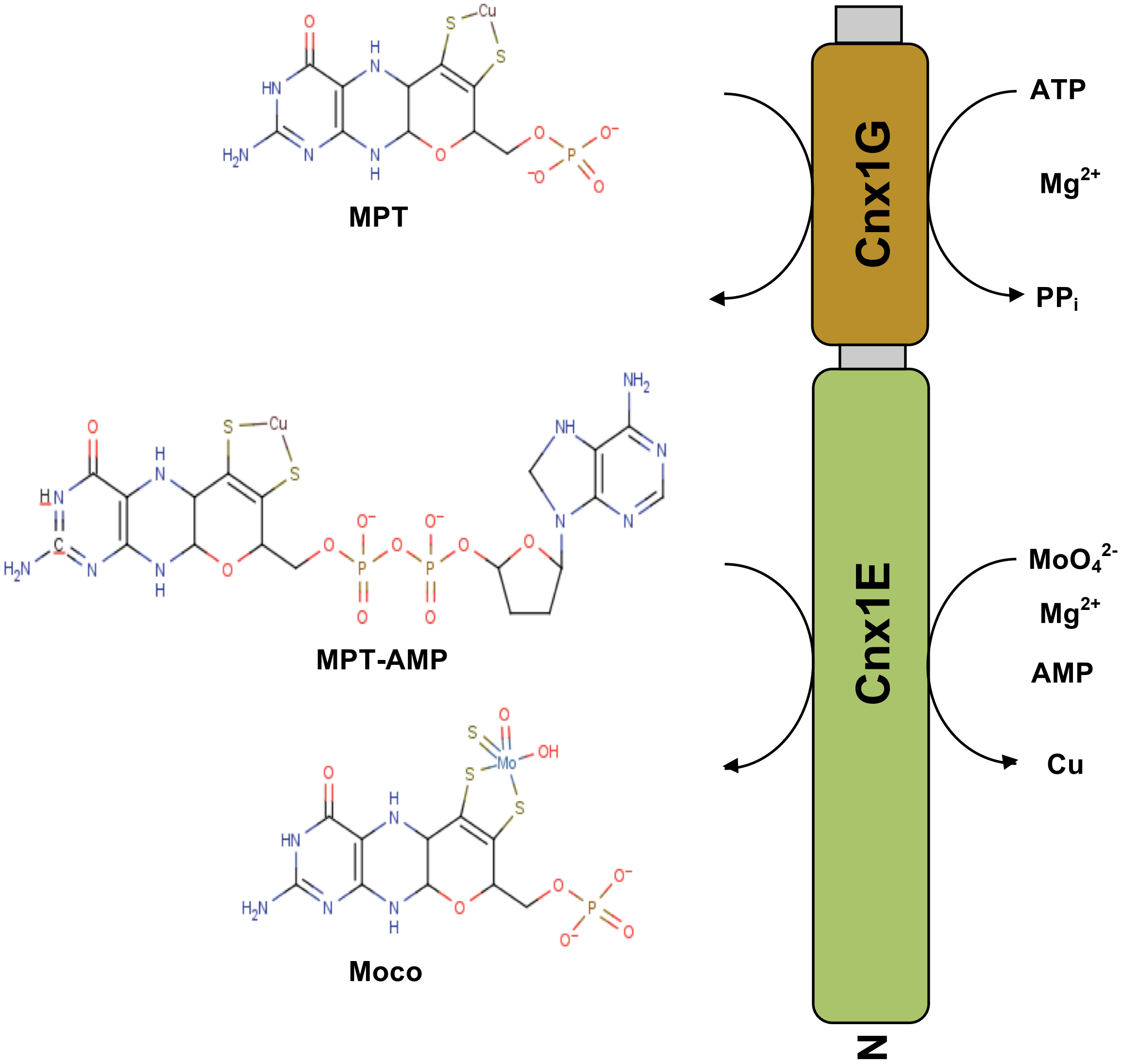


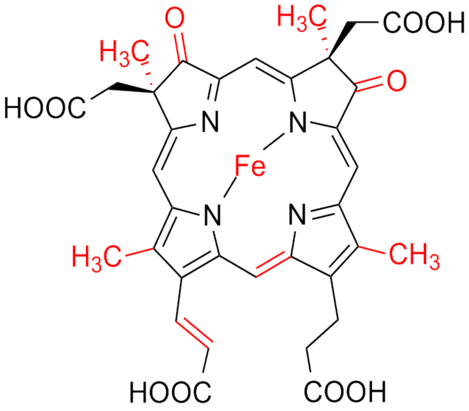
Figure 7: Domain organization of Cnx1. The domain organization of Cnx1 where Cnx1G catalyzes the adenylation of MPT in a Mg^{2+} ATP dependent manner whereas the Cnx1E catalyzes the hydrolysis of MPT-AMP with the subsequent release of copper and insertion of molybdenum. The reaction catalyzed by Cnx1E is dependent on the presence of MoO_4^{2-} and divalent cations like Mg^{2+}.

1A.3. Functional role of Cnx1E

The Cnx1E protein in plants is a 48 kDa protein that is involved in the last and final step of Moco biosynthesis. Previously, it was believed that one domain of Cnx1 is involved in the generation of active molybdenum which is incorporated into MPT by the other domain. However, with the discovery of the novel intermediate MPT-AMP as the product of Cnx1G, the question arose about the functional role of Cnx1E. *In vitro* studies involving the Cnx1G loaded with MPT-AMP demonstrated the synthesis of Moco upon addition of Cnx1E in the presence of Mg^{2+} ions (Kuper *et al.,* 2004). This suggested that Cnx1E is involved in the insertion of molybdenum. Further experiments revealed that in the presence of molybdate, Cnx1E is capable of hydrolyzing MPT-AMP and simultaneously performing metal exchange by releasing the bound copper and inserting molybdenum into MPT, thereby producing mature Moco (Mendel group, unpublished results). However, very little is known about this reaction step, much of which is speculation.

In order to answer the questions regarding this novel mechanism of nucleotide assisted molybdenum insertion, structural information is required which would help to fill the current gap in knowledge. A structure of apo Cnx1E as well as the enzyme-substrate complex of Cnx1E with MPT-AMP is the only means to obtain a detailed structural overview of the active site of the enzyme that would help in identifying the functionally relevant amino acids involved in the catalytic activity. This would ultimately lead to the answers regarding the enzymatic mechanism of Cnx1 and thus provide a detailed understanding of the last two steps of Moco biosynthesis. With this objective, the present work aimed at studying the structure-function relationship of Cnx1E molecule and couple it with the knowledge gained from the structure of Cnx1G (Kuper *et al.,* 2006) to arrive at a definite conclusion on the functional role of Cnx1E.

1B. Role of NirE in the synthesis of the tetrapyrrole heme d_1



1B.1. Tetrapyrroles

As evident from their name, tetrapyrroles are comprised of four pyrrole rings. The pyrrole rings are either linearly attached to one another or arranged in a cyclic fashion in the form of a macrocycle. In the case of porphyrins, they form a macrocycle where the four pyrrole rings are joined by covalent bonds via four unsaturated methine bridges by one-carbon units linking position 2 of one pyrrole ring to position 5 of the next (Figure 8). Tetrapyrroles play a cardinal role in many biological processes such as photosynthesis and respiration. Through various modifications, they form a family of structurally related metallo-prosthetic groups.

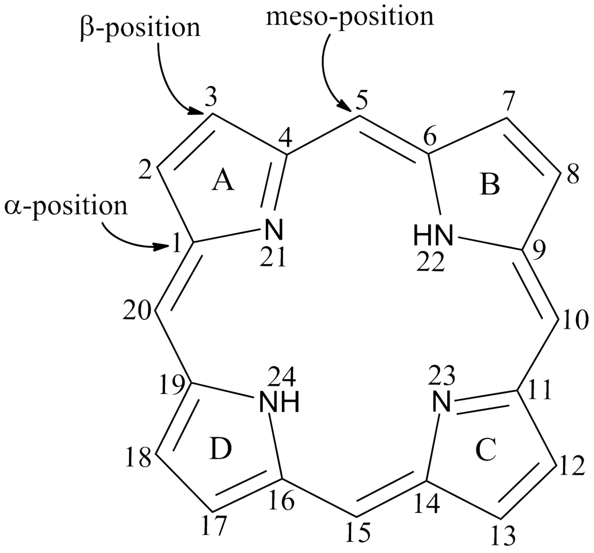


Figure 8: Macrocyclic ring system of porphyrin as the backbone of tetrapyrroles. The pyrrole rings are denoted A-D in a clockwise orientation, carbon and nitrogen atoms are numbered. α-position: carbon atoms next to nitrogen-atoms, β-position: position of substituents, carbon-atoms without direct contact to nitrogens, meso-position: carbon-atoms of methine bridges. (Figure adopted from Sonja Storbeck, *PhD Thesis*, 2011)

Up till now, numerous naturally occurring tetrapyrroles have been deciphered. They can be distinguished from each other based on three characteristics. Firstly, the oxidation states of the macrocyclic ring system; secondly, the nature of the side chains through which they interact with the proteins; and finally, the type of the chelated divalent metal ion e.g. Fe^{2+}, Co^{2+}, Mg^{2+} or Ni^{2+}. This diversity enables them to perform their specific biological function and participate in various enzymatic reactions (Jahn *et al*., 1996). Apart from the macrocycles, linear tetrapyrroles like bilirubin also exist which are derived by the oxidative cleavage of the macrocycle.

Tetrapyrroles play a major role as prosthetic groups of metallo-enzymes, thereby participating in numerous bio-catalytic reactions. Magnesium containing chlorophylls and bacteriochlorophylls convert light to chemical energy during the process of photosynthesis (Allen and Williams, 1998, Beale, 1999). Coenzyme F_{430}, harboring the metal nickel, forms an essential cofactor of methyl-coenzyme M reductase that is involved in the formation of methane (Ermler, 2005, Whitman and Wolfe, 1980). Cobalt containing Cobalamin (vitamin B_{12}) participates in numerous metabolic processes that include methylation, reduction and diverse rearrangements (Martens *et al*., 2002, Raux *et al*., 2000). Linear tetrapyrroles like bilins do not consist of a central metal ion but still play a role in photosynthesis by acting as chromophores (Dammyer and Frankenberg-Dinkel, 2008). Among the iron containing tetrapyrroles, the most important is heme, which is an essential component of hemoglobin that accomplishes the complex task of respiration by its ability to transfer oxygen and exchange gases at cellular level. Siroheme, which also contains iron, is a cofactor of sulfite and nitrite reductases, and is responsible for the catalytic activity of these enzymes in the reduction of sulfite and nitrite (Crane *et al*., 1997). Finally, heme d_1 is another iron-containing tetrapyrrole that is involved in denitrification and is the major subject of this investigation. Figure 9 shows the chemical structures of important naturally occurring tetrapyrroles along with their last common precursor uroporphyrinogen III (uro'gen III).

1B.2. Uroporphyrinogen III: The last common precursor of tetrapyrroles

Tetrapyrrole biosynthesis takes place through a highly conserved, branched, multistep pathway and involves numerous enzymatic reactions including porphyrin ring oxidation, decarboxylation, metal ion chelation and methylation. All naturally occurring tetrapyrroles share their initial steps of synthesis starting with 5-aminolevulinic acid (ALA) as the first precursor (Heinemann *et al*., 2008). ALA is converted to the final common precursor uro'gen III by three enzymatic steps that are common to the synthesis of all tetrapyrroles. From uro'gen III, the tetrapyrrole biosynthesis pathway splits into two, one leading to the intermediate coproporphyrinogen III which is further converted to heme and chlorophyll and the other route leading to the formation of precorrin-2 which concludes in cobalamin, coenzyme F_{430}, siroheme and heme d_1.

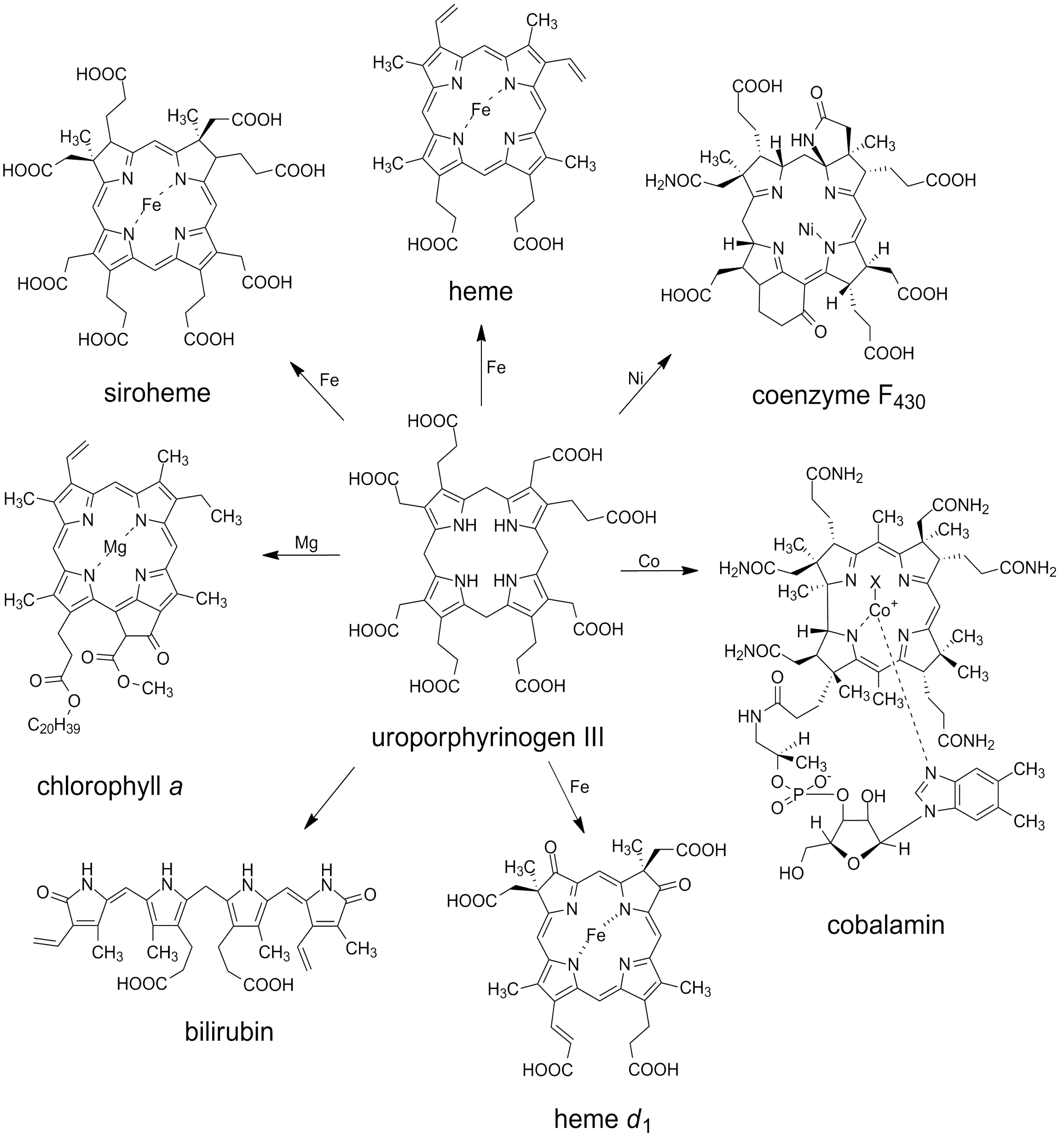


Figure 9: Chemical structures of important representatives of naturally occurring tetrapyrroles. All shown tetrapyrroles share uro'gen III as their last common precursor. The chelated divalent metal ion is denoted next to the arrow. (Figure modified from Sonja Storbeck, *PhD Thesis*, 2011)

1B.3. Heme d1, an essential prosthetic group of the nitrite reductase

Under anaerobic conditions, several species of bacteria such as *Pseudomonas* and *Clostridium* use denitrification as an alternate form of respiration. It consists of four enzyme catalyzed reaction steps via which nitrate is reduced to di-nitrogen. In the first step, nitrate is reduced to nitrite catalyzed by the enzyme nitrate reductase. The nitrite thus formed is converted to nitric oxide in the second step by the enzyme nitrite reductase. The third step involves the reduction of nitric oxide to nitrous oxide by nitric oxide reductase and finally, in the last step, nitrous oxide reductase converts nitrous oxide to molecular nitrogen. The tetrapyrrole heme d_1 serves as an essential prosthetic group of the cytochrome cd_1 nitrite reductase that catalyzes the second step of the denitrification process in denitrifying bacteria such as *Pseudomonas denitrificans*, *Paracoccus pantotrophus* or the human pathogen *Pseudomonas aeruginosa*.

Nitrite reductases are classified into two groups based on their cofactors. One group contains copper as its cofactor and the other having the cofactors heme *c* and heme d_1. In 1991, the crystal structure of a copper containing nitrite reductase from *Achromobacter cycloclastes* was solved, which revealed that the enzyme exists as a homotrimer and contains two types of copper centers (Godden *et al*., 1991). The first type of copper center donates electrons to the second type where the actual catalysis takes place (Ferguson, 1994).

The second group of nitrite reductase contains the cofactors heme ***c*** and heme **d_1** and its members have been appropriately named cytochrome **cd_1** nitrate reductase. The enzyme was first recognized for its oxygen reductase activity (Horio *et al*., 1961) and later for its ability to reduce nitrite (Yamanaka *et al*., 1961). Similar to its counterpart, cytochrome cd_1 nitrate reductase is also a periplasmic enzyme. However, it differs from the copper containing nitrite reductase in its oligomeric assembly, as it exists as a homodimer. Figure 10 shows the crystal structure of the NirS dimer from *Pseudomonas aeruginosa* (PDB ID-1NIR) (Nurizzo *et al*., 1997).

In the dimeric structure of NirS (Figure 10), the N-terminal end of one monomer extends into the active site of the other monomer. Each monomer is composed of two unequally sized sub-domains. The smaller sub-domain harbors a covalently bound heme *c* whereas heme d_1 is non-covalently bound to the larger sub-domain. The substrate binds to the larger sub-domain through the interaction of its nitrogen atom with the ferrous atom of heme d_1 (heme d_1-$Fe^{3+}NO$). The mechanism of this enzymatic reaction involves the transfer of a single electron from heme *c* to the substrate via heme d_1 (Cutruzzola *et al.*, 2003, Einsle and Kroneck, 2004). In the case of NirS, this electron is donated by cytochrome C_{551} (Vijgenboom *et al.*, 1997). An isoelectric nitrosylintermediate, heme d_1-$Fe^{2+}NO^+$ to heme d_1-$Fe^{3+}NO$ was found to exist (Kim and Hollocher, 1983). Heme d_1-$Fe^{3+}NO$ was found to be chemically unstable and releases NO from the active site. Biochemical investigation based on NirS mutants revealed that the anionic ligands like nitrite are stabilized in the active site by two conserved histidine residues that also play a role in the release of the reaction product, nitric oxide (Cutruzzola *et al.*, 2001, Sun *et al.*, 2002). The cofactor heme d_1 plays the cardinal role in this catalytic reaction and forms an essential and unique component of NirS.

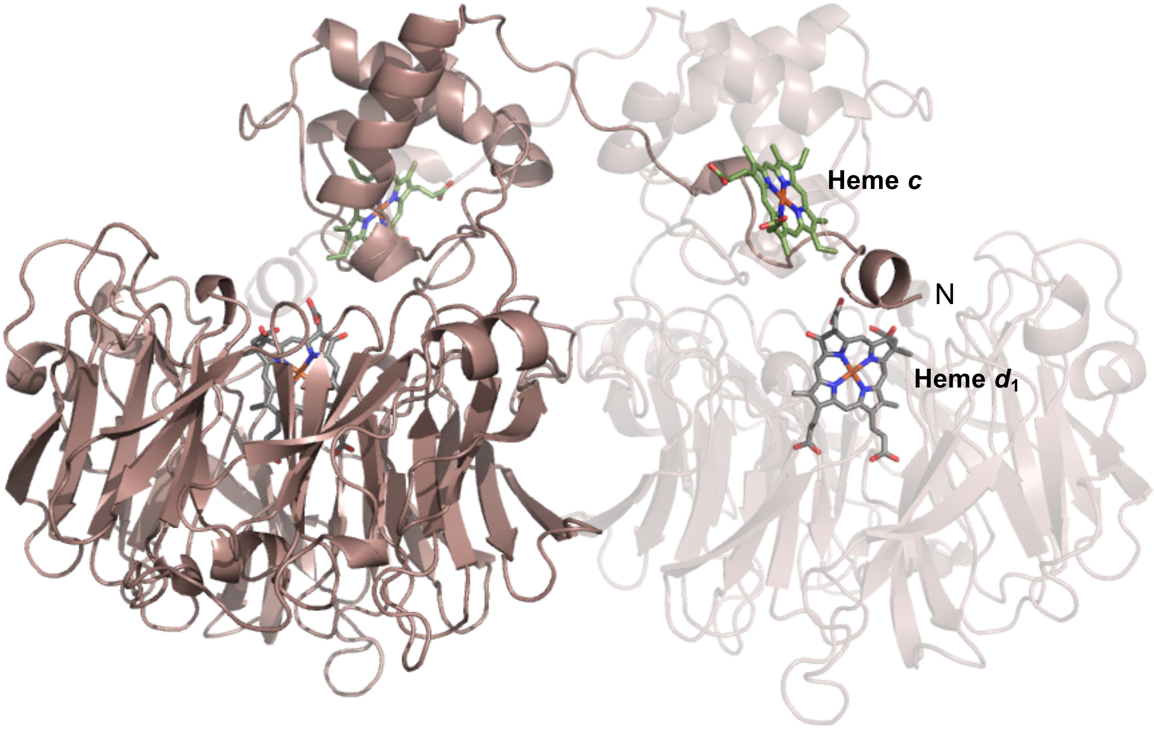


Figure 10: Crystal structure of cytochrome *cd*1 nitrite reductase from *Pseudomonas aeruginosa*. The structure of cytochrome cd_1 nitrite reductase (NirS) is shown in cartoon representation (PDB ID-1NIR). One of the monomer units of dimeric NirS is shown in higher transparency to provide a clear view of the cofactors that are buried deep within the molecule. Each monomer is divided into a small domain harboring heme *c* (green) and a large domain harboring heme d_1 (gray) (Nurizzo *et al.*, 1997).

1B.4. Biosynthesis of Heme d_1

Similar to several other tetrapyrroles like siroheme and cobalamin, synthesis of heme d_1 also follows the route of precorrin-2 from uro'gen III. Previous research based on isotope labeling revealed that the carbon skeleton of heme- d_1 is derived from ALA (Matthews and Timkovich, 1993). Additional experiments involving radioactive labeling of SAM further showed that the methyl groups at rings A and B originate from SAM (Yap-Bondoc*, et al.*, 1990). It has also been proved that the chiral centers at positions C-2 and C-7 possess the same configuration as in siroheme and cobalamin, the latter of which also includes the *cis* configuration of the methyl groups (Aucken *et al*., 1997, Mackman *et al*., 1997, Micklefield *et al*., 1997). Hence, it was concluded that the biosynthetic pathway of heme d_1 coincides with several reaction steps of siroheme and cobalamin biosynthesis. The route of heme d_1 biosynthesis can be summarized as follows: uro'gen III is first methylated at positions C-2 and C-7, followed by the oxidation of the macrocycle and insertion of a divalent iron. Then, the acetate groups at positions C-12 and C-18 are decarboxylated, the propionate groups are removed from C-3 and C-8 and oxo-groups are introduced at these positions. Finally, an acrylate side chain is formed at C-17, thus completing the synthesis of heme d_1.

In *Pseudomonas aeruginosa*, a set of genes known as the *nir* genes have been identified, many of which encode enzymes involved in the biosynthesis of heme d_1. Figure 11 shows the organization of *nir* genes in the *nir*-operon as suggested by Kawasaki (Kawasaki *et al*., 1997). The first gene, *nirS,* encodes the cytochrome cd_1 nitrite reductase NirS. Based on mutational analysis, it was proposed that the genes *nirF-E* encode enzymes that are responsible for heme d_1 biosynthesis (Kawasaki *et al*., 1997).

The protein encoded by *nirF* was found to exhibit 35% sequence identity to NirS (Bali *et al*., 2010, Timkovich, 2002). This protein was also found to possess a lysine rich motif that could be involved in the binding of nucleotides (Scrutton *et al*., 1990). Additionally, it was also seen that in *Paracoccus pantotrophus,* NirF is a periplasmic protein carrying an N-terminal signal sequence for Sec dependent translocation. In *Pseudomonas aeruginosa*, NirF was found to possess a signal sequence generally associated with the liproproteins of the cytoplasmic membrane (Sonja Storbeck, *PhD Thesis*, 2011). On the basis of these results, it was proposed that NirF is a periplasmic protein with a functional role during the late steps of heme d_1 synthesis (Bali *et al.*, 2010).

The genes *nirD, nirL, nirG* and *nirH* encode proteins that show pair wise sequence identity, with NirD and NirG having 45% and NirL and NirH 35% identity. It was suggested that these enzymes form a heterotetrameric complex (Timkovich, 2002), which was also seen by protein interaction studies (Sonja Storbeck, *PhD Thesis*, 2011). These four proteins were found to play the role of transcriptional regulators of the *nirJEN* genes (Sonja Storbeck, *PhD Thesis*, 2011). In addition, these proteins have also been shown to catalyze the decarboxylation of the acetate groups at positions C-12 and C-18 of siroheme to corresponding methyl groups (Bali *et al.,* 2011).

The enzyme encoded by the next gene of the operon, *nirJ*, belongs to the radical SAM protein family (Brindley *et al.*, 2010, Sofia *et al.*, 2001). It harbors the typical CXXXCXXC consensus sequence required for the coordination of a [4Fe-4S] cluster that aids in the conversion of SAM to 5'-deoxyadenosyl radical by furnishing an electron (Wang and Frey, 2007). Such radical based reaction mechanisms are characteristic of challenging chemical transformations. NirJ is believed to be involved in attaching the oxo-groups at positions C-3 and C-8 (Zumft, 1997).

Finally, the *nirE* gene encodes a SAM dependent uro'gen III methyltransferase (SUMT), which was proposed to be involved in the methylation of uro'gen III at positions C-2 and C-7 to yield precorrin-2. This hypothesis was based on the high level of sequence identity of NirE with CobA from *Pseudomonas denitrificans* and CysG from *E. coli* and *Salmonella typhimurium* (Timkovich, 2002).

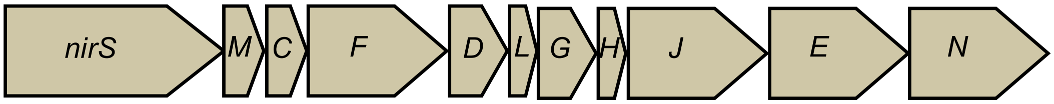


Figure 11: The *nir* genes as organized in the *nir*-operon of *Pseudomonas aeruginosa*. The brown arrows represent the genes of the *nir*-operon with the corresponding name of the gene denoted within.

1B.4.1. Working model for heme d_1 biosynthesis

On the basis of the above hypothesis of the functional role of enzymes encoded by the *nir* operon, a model for the biosynthesis of heme d_1 can be proposed (Figure 12). In this model, heme d_1 biosynthesis starts with ALA that is converted to uro'gen III by the enzymes HemB, HemC and HemD. From this point, the heme d_1 synthesis route branches off. At first, uro'gen III is converted to precorrin-2 in a SAM dependent methylation reaction catalyzed by NirE. Following this, precorrin-2 is oxidized by a dehydrogenase to yield sirohydrochlorin, which is then acted upon by a ferrochelatase to insert an iron, thus converting sirohydrochlorin to siroheme. Currently, the source of the dehydrogenase and ferrochelatase is not known. In the next step, siroheme is converted to 12,18-di-decarboxysiroheme by a decarboxylation reaction possibly catalyzed by the tetrameric complex of NirDLGH. This is followed by the addition of two oxo-groups at positions C-3 and C-8 by the enzyme NirJ, thereby yielding 3,8-dioxo-12,18-dicarboxy-siroheme. Finally, an acrylate side chain is formed at C-17 by the enzyme NirF, thereby producing heme d_1.

The NirE catalyzed conversion of uro’gen III to precorrin-2 in a SAM dependent methylation reaction is one of the central themes of this investigation and has been studied in detail through this work. An attempt was made to understand the mechanism of this reaction from a structural aspect of the enzyme, NirE. This has been described further in this thesis.

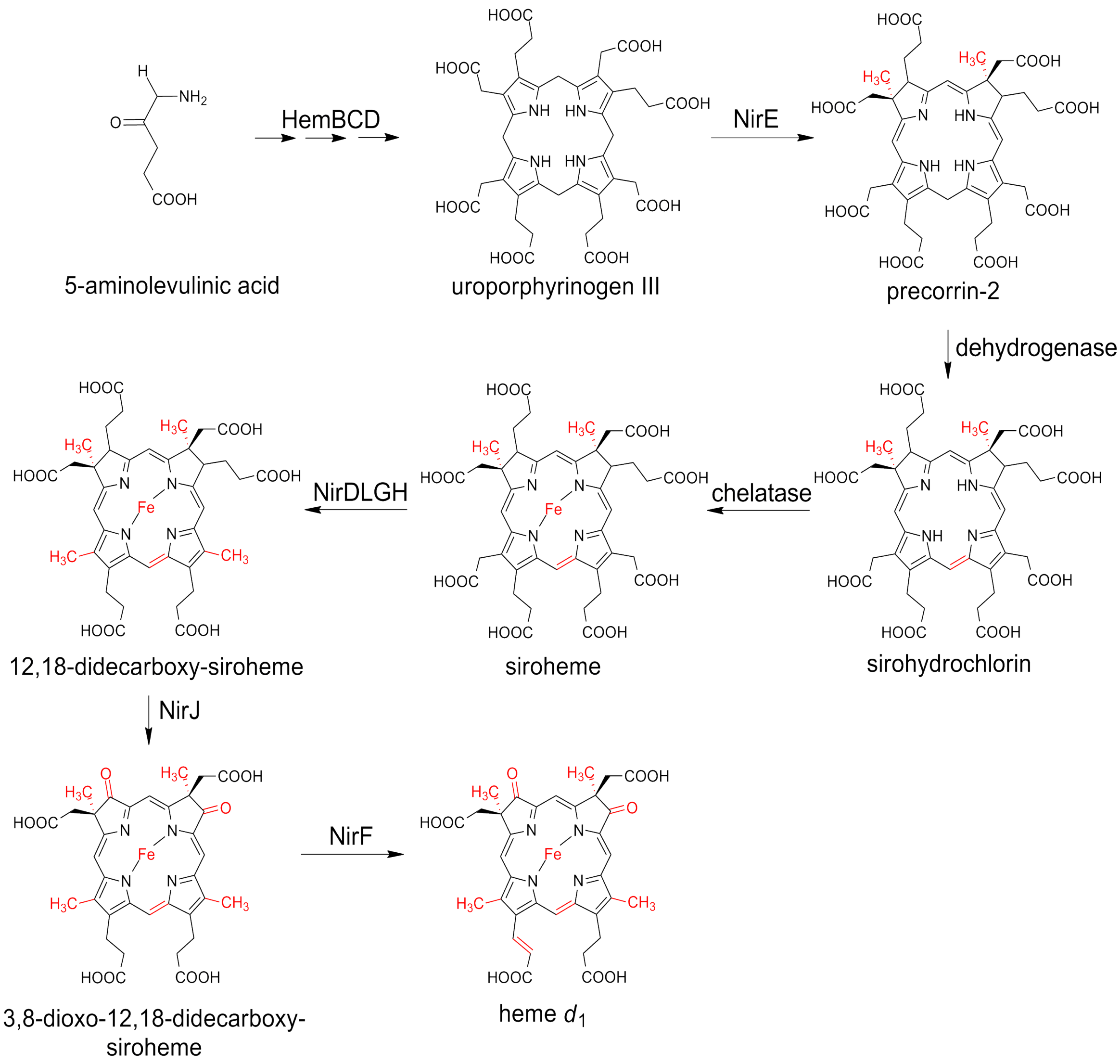
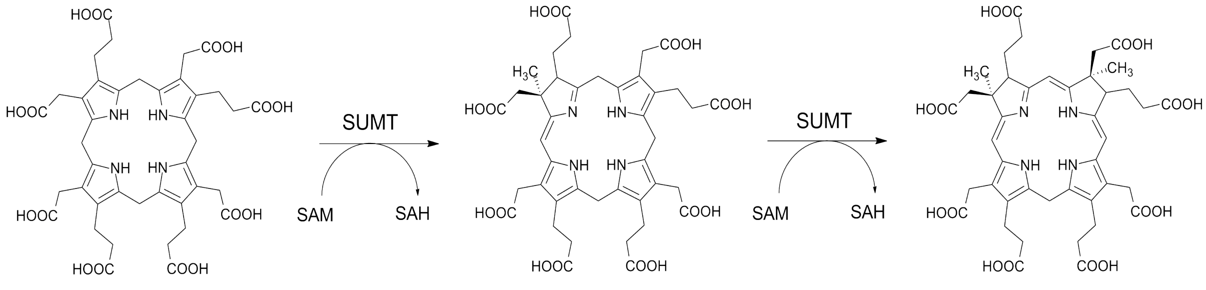


Figure 12: Working model for heme *d*1 biosynthesis. The working model for heme *d*1 biosynthesis shows the conversion of uro'gen III to precorrin-2 by NirE. Further steps lead to the oxidation of precorrin-2 to sirohydrochlorin and iron insertion to yield siroheme. It is presumed that NirDLGH catalyzes the decarboxylation of siroheme resulting in the formation of 12,18-didecarboxy-siroheme. Following this, NirJ generates the oxo-groups through a potential radical mechanism. Finally, the formation of an acrylate by NirF in a dehydrogenation reaction completes the synthesis. (Figure adopted from Sonja Storbeck, *PhD Thesis,* 2011)

1B.4.2. SAM dependent methylation of uro'gen III

Heme *d_1* biosynthesis, as discussed in the previous section, proceeds through precorrin-2, produced by the methylation of uro'gen III at positions C-2 and C-7. During the biosynthesis of cobalamin and siroheme, the methyl groups are furnished by two SAM molecules and the methylation is catalyzed by a family of enzymes known as SAM dependent uro'gen III methyltransferases (SUMT). It is a two step reaction with the methylations occurring sequentially (Figure 13). The first methylation occurs at C-2 yielding the intermediate precorrin-1, which is further methylated by a second SAM molecule at position C-7. Following each methylation, the SAM molecule gets converted to S-adenosyl-L-homocysteine (SAH) and leaves the active site. The methylation takes place by a S_N2 mechanism (Mascaro *et al*., 1977) and each methylation is presumed to be followed by a prototropic rearrangement of macrocycle double bonds (Stroupe *et al*., 2003, Warren *et al*., 1994).



Uroporphyrinogen III **Precorrin-1** **Precorrin-2**

Figure 13: SAM-dependent methylation of uro'gen III leading to the formation of precorrin-2. The conversion of uro'gen III to precorrin-2 takes place by a two step reaction with the formation of the monomethylated intermediate precorrin-1. (Figure adopted from Sonja Storbeck, *PhD Thesis,* 2011)

Up until now, a number of SUMTs have been discovered from various organisms and their roles in the biosynthesis of cobalamin (CobA), siroheme (CysG, SirA, UPM1, Met1p) and heme d_1 (NirE) have been studied and biochemically characterized. Based on their molecular size and enzymatic activity, these SUMTs can be classified into three subclasses. The first subclass consists of SUMTs like CobA, SirA, Met1p, and NirE. These are monofunctional enzymes exhibiting only SUMT activity. These proteins are generally dimeric molecules with each subunit having a molecular weight of roughly 30 kDa. The second subclass includes the protein CysG which is also a dimeric protein but of higher molecular weight, each of its subunits measuring 50 kDa. The C-terminal domain of CysG is homologous to CobA, SirA, Met1p, and NirE and exhibits SUMT activity. The N-terminal domain has the dual activity of dehydrogenase-ferrochelatase and is responsible for the conversion of precorrin-2 to siroheme via the intermediate sirohydrochlorin. Thus, the second subclass consists of a trifunctional member with higher molecular weight. The third subclass comprises of fusion proteins, where a SUMT belonging to the first subclass, such as CobA, is linked to HemD, giving rise to a protein having the dual function of uro’gen III synthase along with SUMT activity.

Several crystal structures of SUMTs have been solved which include CobA from *Pseudomonas denitrificans* (Vevodova *et al*., 2004), CysG from *Salmonella enterica* (Stroupe *et al*., 2003) and the SUMT from *Thermus thermophilus* (Rehse *et al.*, 2005). These structures were solved with the SAH bound to the active site. Based on the structural information obtained from these SUMTs, mutational analysis has been performed that helped in indentifying the amino acids that are involved in the coordination of SAH and SAM. Additionally, on the basis of *in silico* docking of uro'gen III at the enzyme active site, it has been proposed that uro'gen III is bound above SAH. Docking also indicated the amino acids that could be involved in the binding of uro'gen III and possibly in catalysis. However, due to the lack of an enzyme-substrate structure, these speculations could not be verified. In order to gain a complete understanding of the reaction mechanism of SUMTs, it is essential to obtain a structure of a SUMT in complex with its substrate uro’gen III. This would give a structural insight into the active site of the enzyme and also help in identifying the functionally important amino acids. The structure of the enzyme-substrate complex would help in reaching a concrete conclusion regarding the enzymatic mechanism of SUMTs. This work aims to address these questions and understand the detailed mechanism of SUMT enzymes from a structural point of view.

2. Aims and Scopes

The central objective of this work was to elucidate the mechanism of the chemical reaction catalyzed by the enzymes Cnx1E and NirE. Both the enzymes have been implicated in the conserved biosynthetic pathway leading to the synthesis of their respective cofactors. A complete understanding of their catalytic mechanism can only be reached through the structure of the enzyme-substrate complex of both the proteins.

In order to achieve this goal, both Cnx1E and NirE need to be produced recombinant. The recombinant protein was then to be used for obtaining protein crystals and for biochemical characterization by functional assays. On obtaining single diffracting quality protein crystals, the three dimensional structure of the molecule was to be solved by X-ray crystallography. X-ray crystallographic analysis would provide a detailed structural insight into the active site of the enzymes. Based on the structural information obtained from the enzyme-substrate complex, the putative amino acids that could be involved in catalysis were to be identified. In the event of the absence of an enzyme-substrate complex as in the case of Cnx1E, the substrate was to be modeled into the putative active site to obtain an idea regarding the functionally relevant amino acids. Further to this, site directed mutagenesis of the identified amino acids was to be performed to obtain the protein variants, which should be then subjected to functional characterization by enzyme assays. This would enable in identifying the residues involved in substrate coordination as well as the catalytically essential residues and ultimately establish the structure-function relationship of the protein molecules. Finally, this would provide a conclusive understanding of the enzymatic mechanisms of Cnx1E and NirE.

3. Materials and Methods

3.1. Protein Production

3.1.1. NirE

The expression and purification of NirE was done by Sonja Storbeck as a part of her Ph.D. thesis. NirE was expressed and purified as described previously (Storbeck *et al.,* 2009, Storbeck *et al.,* 2011). Briefly, the protein was expressed in *E. coli* BL21(DE3) chemically competent cells. Upon protein expression the cells were harvested and resuspended in buffer A [50 mm Tris ⁄ HCl (pH 7.5), 200 mm KCl, 10% (w ⁄ v) glycerol]. Cell lysis was performed using a French press at 1000 p.s.i. (68947.57 hPa) and the soluble protein was separated from the cell debris by centrifugation for 1 hour at 175 000 g. The NirE protein contained in the supernatent was purified with a 1.5 mL Ni^{2+} SepharoseTM 6 Fast Flow column equilibrated with buffer A. After applying the protein to the column, it was washed several times with buffer A and finally eluted with 2.5 mL of buffer A containing 200 mM imidazole. The eluted fractions containing NirE were pooled and dithiothreitol (DTT) was added to a final concentration of 5mM. PreScissionTM Protease or Tag off enterokinase (Novagen) was added to the NirE solution which was dialyzed over night against buffer A. Following dialysis, the tag was removed by applying the protein solution to Ni^{2+}-Sepharose^{TM} 6 Fast Flow column. Finally the untagged protein was further purified by size exclusion chromatography using a HiLoad 16 ⁄ 60 Superdex 200 column (GE Healthcare) equilibrated with buffer A. The NirE containing fractions were pooled and concentrated by ultrafiltration (Amicon, Millipore GmbH, Eschborn, Germany) and subjected to buffer exchange in an anaerobic chamber (Coy Laboratories, Grass Lake, MI, USA) against buffer A. Purified NirE was directly used for setting up crystallization.

3.1.2. Cnx1E:

Cnx1E and its variants were expressed and purified by Corinna Probst as a part of her Master's thesis (Corinna Probst, Master's thesis 2012). Briefly, the Cnx1E protein was expressed as a streptavidin tagged fusion protein in *E.coli* RK5204 chemically competent cells. After protein expression, the cells were harvested by centrifuging at 17000 g. The cell pellet was lysed using a French press at 1000 p.s.i. the soluble protein was separated from cell debris by centrifugation and the supernatant containging Cnx1E was subjected to purification using 3 mL StrepTactin-Matrix (High Cap StrepTactin-Matrix, IBA).

3.2 Functional assays

3.2.1. NirE

3.2.1.1. Determination of NirE activity

The determination of the catalytic activity of NirE wildtype and NirE mutants was done in an established in vitro activity assay (Strobeck *et al*., 2009). The activity assay was performed under anaerobic conditions in an anaerobic chamber (Coy Laboratories). 1 mM 5-aminolevulinic was used as the starting compound to synthesize Uro'gen III using the enzymes HemB from *P. aeruginosa* (0.14 µM), HemC (0.15 µM) and HemD (0.17 µM) both from *B. megateriumin* in a final volume of 1 mL of thoroughly degassed buffer containing 50 mM Tris-HCl pH 8.0, 100 mM KCl, 5 mM $MgCl_2$, 50 mM NaCl, 5 mM DTT. The final concentration of NirE wildtype and NirE mutants used was 1.5 µM. For quantification purpose, precorrin-2 was converted to sirohydrochlorin by the addition of SirC from *B. megaterium* (1.5 µM) and NAD^+ (100 µM). Quantification was done using the published extinction coefficient of sirohydrochlorin (ε376nm = 2.4 x 10 5$M^{-1}cm^{-1}$, (Leech *et al*., 2003)).The reaction was started by adding 200 µM of SAM. The assay mixture was incubated at 37 °C in the dark and product formation was monitored using a V-550 or V-650 spectrophotometer (Jasco, Groβ-Umstadt, Germany). The determination of specific activity of NirE was done using chemically reduced uro'gen III at a final concentration of 17 µM.

3.2.1.2. SAM binding assay

The SAM-binding assay was performed by an already established protocol (Woodcock *et al*., 1996). Purified NirE protein at a concentration of 100 µM was incubated with 0.5 µ Ci of [methyl-^{14}C]-S-adenosyl-l-methionine in a final volume of 250 µL of buffer A at 25 °C for 1 hour. Following the 1 hour incubation, the protein solution was then passed over a NAP-5 column (GE Healthcare) and eluted with 3 mL of buffer A.100 µL fractions of were collected and analyzed for radioactivity using a Liquid Scintillation Analyzer Tri-

Carb 2900TR (PerkinElmer Life Sciences). The binding of SAM was analyzed by fluorography (Woodcock *et al*., 1996).

3.2.2. Cnx1E

3.2.2.1. HPLC-based detection of intermediates of the Moco biosynthesis

Moco and its precursor MPT and MPT-AMP are very unstable and extremely sensitive to oxidation. Therefore, they degrade very quickly to their oxidation products FormA and FormB, as well as non detectable by-products. These degraded oxidation products fluoresce due of their conjugated π-electron system. Hence it is possible to quantitatively detect Moco and its precursors by using fluorescence measurement by HPLC (high pressure liquid chromatography) on these oxidation products. The formation of only one oxidation product (FormA-AMP) is enforced by applying extremely acidic and oxidizing conditions, which makes the detection much easier. Quantification of MPT / Moco and MPT-AMP is in this work is based on the FormA-peak. The area under this peak corresponds to the total amount of FormA and is calculated by integration. The following formula is employed to calculate MPT/Moco- or MPT-AMP-saturation:

$$n(MPT/Moco) = A_{FormA\ of\ MPT\ bzw.MOCO} \times \left(\frac{total\ volume}{analyzed\ volume}\right)$$

$$n(MPT - AMP) = A_{FormA\ of\ MPT-AMP} \times \left(\frac{total\ volume}{analyzed\ volume}\right) \times 1{,}6$$

n = Stoffmenge [mol]

A = Flächeneinheit (Integral des Peaks)

3.2.2.2. Oxidation and dephosphorylation of Moco / MPT

The protein solution to be analyzed (normally 500 pmol protein) was mixed with 0.1M Tris-HCl pH 7.2 and made up to a volume of 400 µL and 50 µL fresh oxidation solution (1% I_2/KI solution and 86 µL 37% HCl) was added.The oxidation was carried out in the dark at room temperature overnight. Following the oxidation step, the precipitated protein and other deposits are pelleted by centrifugation for 10 minutes at 11000 g.

The supernatant was transferred to a new tube and mixed with 56 µL of a freshly prepared 1% solution of ascorbic acid to reduce the excess iodine. The pH was raised to about 8.3 by adding 200 µL of 1M Tris (unbuffered), so that the subsequent alkaline phosphatase reaction could be performed. In addition, 13 µL of 1M $MgCl_2$ was added. 2 µL alkaline phosphatase (AP, 1U/µL) was added to dephosphorylate FormA which allows better resolution in HPLC-based FormA analysis. The reaction mix was incubated at room temperature at room temperature overnight after which it was treated with AP and centrifuged for 10 min at 11000 g and the supernatant is removed. Upon dephosphorylation, dephospho-FormA is purified by isocratic anion exchange chromatography.

3.2.2.3. Isocratic anion exchange chromatography

Isocratic anion exchange chromatography enables the purification of the obtained dephospho-FormA. For this purpose, a QAE-Sephadex A 25 matrix (Sigma) was used, which acts as a strong anion exchanger. The matrix was equilibrated overnight in 1M NaCl and 1M ammonium acetate buffer in order to convert it to its acetate form. After the equilibration step, 0.5 mL of the matrix material was filled in empty NICK columns and sealed by putting a watered frit on top of the matrix which is then used for purification.

Prior to purification, the matrix is soaked 2 times with 1M ammonium acetate in order to remove any residues and convert the matrix completely to its acetate form. The column is rinsed twice with filtered distilled water and the centrifuged sample is added to the column. After loading the sample, the column is again washed with distilled water to remove unbound impurities. The elution of dephospho-FormA is performed using 10 mM acetic acid. Initially, a dead volume is added to the column of 300 µL. The bound dephospho-FormA was eluted with 10 mM acetic acid 800µL. The sample is now ready for the HPLC analysis, but it can also be stored at -20 ° C.In order to remove the dephospho-FormA which is still bound to the column, it was again washed with 3 mL of 10 mM acetic acid. The elution of FormA-AMP was done by means of 50 mM HCl. First, a volume of 300 µL is added to the column and allowed to sink in. Then, the bound FormA-AMP is eluted with 800 µL 50mM HCl.

3.2.2.4. FormA-AMP-determination

In order to perform the HPLC analysis of the eluted FormA-AMP, it was first converted to FormA.

The FormA-AMP-elution was divided into two aliquots of 350 µL and 25.2 µL. The pH of each aliquot was raised to 7.5 by adding 1M Tris (unbuffered). 5.25 µL of $MgCl_2$ (1M) and 1 µL phosphodiesterase I (1U/µL) was added to the aliquots to de-adenylate FormA-AMP. The reaction mix was incubated overnight in the dark at room temperature. In order to stop the reaction, the aliquots were incubated for 5 min at 96 °C and then incubated on ice for 3 min. The dephosphorylation was carried out in the dark, overnight, at room temperature by adding 10.5 µL of 1M Tris (unbuffered) and 2 µL alkaline phosphatase (1U/µL). Both aliquots were combined. Then 500 µL of the combined sample are analyzed together with 300µL of 10 mM acetic acid by HPLC.

3.2.2 5. HPLC analysis of dephospho-FormA

The eluates obtained are analyzed by HPLC chromatography on a C18 reverse phase column and detected by their fluorescence. The fluorescence detector (G1321A, Agilent) of the HPLC pump (G1312A Bin pump, Agilent) was used and was operated with the highest sensitivity (gain 16x). The fluorescence of FormA is measured after 370 nm excitation and 450 nm emissions. The flow rate was maintained at 1 mL/min. The running buffer, a mixture of 5 mM NH_4Ac and 15% methanol was chosen. The proportion of methanol in isocratic flow determines the retention time of FormA-dephospho. The detected fluorescence was recorded using the Agilent ChemStation software, and the area of the FormA-peak was determined.

3.2.2.6. *In-vitro* MPT-AMP hydrolysis at Cnx1E-MPT-AMP complex

The formation of physiologically active Moco by Cnx1E is dependent on divalent cations (Llamas *et al*., 2006). After formation of the Cnx1E-MPT-AMP complex, MPT-AMP hydrolysis and the incorporation of molybdenum take place simultaneously to yield physiologically active Moco.

Hydrolysis is induced by adding $MgCl_2$ to 500 pmol of Cnx1E MPT-AMP complex to a final concentration of 5 mM. The reaction mix was incubated at room temperature for 20 minutes.

The reaction is stopped by the addition of oxidizing solution. Moco and its precursors were detected by the same method as HPLC based formA detection.

3.2.2.7. Colorimetric detection of molybdate in biological materials and solutions

For quantitative detection of molybdate in biological materials and solutions a photometric method was used. The detection method is based on the formation of a green molybdate dithiol complex that is formed by the addition of toluene-3,4-dithiol under acidic conditions. This complex can then be extracted and detected spectrophotometrically at a wavelength of 680nm.

The protein solution to be analyzed was made up to a final volume of 100µL with a buffer containing 0.1M Tris/HCl pH 8.0, 125mM NaCl and 10% glycerol and denatured by heating to 96 °C for 30 minutes. Following this, the mixture was centrifuged for 10 minutes at 14000 g at room temperature in order to pellet precipitated protein, and the supernatant was transferred to a 500µL reaction tube. The complex formation was started by adding 200 µL of the dithiol reagent and H_2SO_4 to a final concentration of 8N and incubated for 15 minutes at room temperature. 100 µL isoamylacetate may be added to the mixture and vortexed for 1minute to extract the colored complex in the organic phase. To speed up the phase separation, the mixture was centrifuged at 8000 g for 10 minute at room temperature. 80 µL of the upper phase was removed and measured at 680nm against an isoamylacetate blank.

3.2.2.8. Calibration line for the determination of molybdate concentration

A calibration curve is computed by linear regression from the absorption of molybdate solutions of different concentrations at 680nm. With help of this curve, the specific molybdate saturation of the protein solutions (pmol molybdate/pmol protein) was determined.

3.3.Crystallographic Methods

3.3.1. Protein crystallization

Prior to setting up crystallization trials, purified and concentrated protein was centrifuged at 18.800 g for 20 minutes at 4 °C. Initial screening was attempted by sitting drop vapor diffusion method using an automated crystallization robot. After obtaining initials hits, further optimization was performed manually by hanging drop vapor diffusion method.

3.3.1.1. Crystallization of NirE

For crystallizing NirE in the absence of its substrate uro’gen III, the enzyme at a concentration of 9.2mg/mL was incubated with 2mM SAM for 1 hour prior to use. Crystallization trials were done in sitting drop vapor diffusion set ups using an automated Mosquito crystallization robot TTP Lab Tech Ltd., Melbourne, UK). For the crystallization drops, 200nL of NirE solution was mixed with 200 nL of buffer solution from JCSG I-IV and ammonium sulfate screens (Qiagen, Hilden, Germany). Crystals grew within 3-6 days and were soaked with 20 % glycerol prior to shock cooling in liquid nitrogen.

For crystallization of NirE along with its substrate uro’gen III the protein was used at a concentration of 10 mg/mL. Uro’gen III was obtained by reducing uroporphyrin III by chemical reduction using H_2/palladium as previously described (Layer *et al*., 2002). Crystallization was set up in an anaerobic chamber at 17 °C by sitting drop vapor diffusion method. Prior to crystallization, the protein was incubated with an 82-fold molar excess of uro’gen III and 1 mM SAH for a period of 1 hour. 1 µL of protein solution was mixed with 1 µL of reservoir solution of the JCSG+ Screen (Qiagen, Hilden, Germany) and incubated in the dark. Crystals grew within 4-8 days.

3.3.1.2. Crystallization of Cnx1E

Initial screening

Initial crystallization trials were done with commercially available screens from Qiagen (see Chapter 3.1.3) in 96-well sitting drop vapor diffusion plate (Greiner). 200nL of pure protein at a concentration of 10mg/mL was mixed with an equal volume of reservoir solution from the commercial screens using the Honeybee 961 crystallization robot (Zinsser Analytics). The plates were sealed using a MancoTM Crystal Clear tape (Jena Bioscience) and incubated at 18 °C which facilitated the equilibration of the drops against the reservoir solution. The plates were checked at an interval of 3 days for the appearance of crystals with a microscope.

Optimization

Initial crystals obtained from the 96 well plates were tested for possible diffraction and further optimized in using 24 well hanging drop vapor diffusion plates. The main parameters that were varied for optimization were protein and precipitant concentrations and pH of the buffer. The drop size was also increased to 4 µL. Crystallization was set up using 2 µL of protein mixed with an equal volume of reservoir solution. Additionally, seeding techniques like streak, micro and micro-seeding (Bergfors, 2003; Luft & DeTitta, 1997), Additive Screen (Hampton Research) were also tried to obtain larger crystals with better diffraction properties.

3.3.2. Cryo protection

Protein crystals upon irradiation by x-rays result in the generation of free radicals within them. These radicals can migrate through the crystals and damage them, thereby resulting in rapid loss of diffraction. To reduce this damage, crystals can be subjected to shock freezing to 100K in liquid nitrogen. However, aqueous solutions on being frozen can lead to the formation of ice which also have the potential to damage the crystal and compromises the quality of the measured data due to its own diffraction (ice rings). In order to overcome this problem, cryo protectants such as small molecular weight PEG, glycerol or sugars can be added. Cryo protectants were prepared by mixing 10 – 20 % of glycerol to reservoir buffer. Crystals were removed from the crystallization drop by a nylon loop to this solution of cryo protectant. Following this the crystals were quickly frozen in liquid nitrogen.

3.3.3. Data collection

The crystals used for data collection were shock frozen in liquid nitrogen and were maintained at 100 K by a continuous stream of cold nitrogen gas. Crystals were screened in house using a Rigaku MicroMax 007HF rotating anode x-ray generator rotating anode x-ray generator at a wave length of 1.54 Å with a Saturn Image plate detector. Complete data sets for NirE crystals were collected at the beam lines of Berliner Synchrotron Gesellschaft m.b.H (BESSY) in Berlin, Germany with a Raynonix MX225 3x3 CCD detector (Marresearch GmbH Norderstedt, Germany) at a wave length of 0.9537Å. For the Cnx1E crystals, the data collection was done in house using the same x-ray source and detector as used for screening.

Collection of diffraction data was done by the oscillation method which implies the rotation of the crystal in small oscillation steps (Φ) around an axis perpendicular to the incident x-ray beam. Crystal mosaicity and cell parameters are the main determinants on which determines the optimal oscillation range for each image; for the experiments carried out in this investigation, the typical value was 0.8°. As a result of each oscillation, a diffraction image was generated which exhibited a pattern of strong reflections (i.e. BRAGG peaks). The intensities of these peaks and their spatial arrangement reflects the electron density function ρ(x,y,z) of the crystal. BRAGG peaks emerge as a result of the elastic scattering of incident beams by the atoms in the crystal and the constructive interference of these scattered waves, which takes place when their phase shift is a multiple of 2π. This condition can be expressed by the BRAGG equation.

$$n\lambda = 2d \sin\theta \qquad \text{(equation - 1)}$$

The EWALD sphere can be used to illustrate the implications of the equation-1. The EWALD sphere is a geometric construct in the reciprocal space whose radius is equal to λ^{-1}. A crystal can be regarded as a lattice of points of equal symmetry defined by its primitive vectors a, b, c. A related reciprocal lattice with the respective vectors a^*, b^*, c^* can be generated in reciprocal space. The oscillation in real space corresponds to a rotation of the reciprocal lattice through the EWALD sphere. The BRAGG diffraction condition as expressed in equation-1 is satisfied when the surface of the EWALD sphere intersects a grid point of reciprocal lattice and these results in an emerging beam which generates a reflection on the surface of the detector. For collecting a complete data set, crystallographic data needs to be collected over an angular range which is dependent on the orientation of the crystal as well as its inherent symmetry.

In this investigation, XDS (Kabsch 2010) and HKL suite (Minor *et al.,* 2002) was used to determine the crystal parameters, assign space group, index and integrate the images' reflections. The scaling of intensities and the merging of partial reflections were done with SCALA (Evans, 2006) of the CCP4 suite (CCP4, 1994) and converted to structure factor amplitudes with TRUNCATE. The R_{sym} value which can be calculated for the resolution shells, determines the quality of a data set. It shows how well intensities of the symmetry related reflections hkl are correlated. The R_{sym} is calculated by the following formula:

$$R_{sym}\ (I)\ = \frac{\sum_{hkl} \sum_{i} \left| I_i\,(hkl) - \overline{I\,(hkl)} \right|}{\sum_{hkl} \sum_{i} I_i\,(hkl)} \qquad \text{(equation - 2)}$$

Where I (hkl) is the average intensity and I_i (hkl) is the scaled intensity of the observations of a given reflection. A high R_{sym} must be avoided by scaling the data set to a lower high resolution limit. Another quality control criteria which was monitored was the value of I/σ(I). When the value of I/σ (I) was less than 2 scaling of the data was repeated with a lower high-resolution limit as it was considered to be an extremely low signal-to-noise ratio. Finally the completeness of the data set is calculated by the following formula:

$$C = \frac{n_{\text{obs}}}{n_{\text{theo}}} \qquad \text{(equation - 3)}$$

n_{obs} is the number of observed reflections and n_{theo} is the number of theoretically observable reflections for the resolution shells. When the completeness in a shell is less than 0.95, it indicates a lack of data for this resolution range which can be either avoided by scaling to a tighter resolution limit (in case the highest resolution shell in incomplete) or by increasing the total angular range of the data collection.

3.3.4. Phasing by Molecular Replacement

Similar to all electromagnetic waves, the three parameters frequency, amplitude and phase characterize an x-ray scattered by the atoms of a crystal. The intensity and the position of the BRAGG peaks determine the amplitude and frequency of the emerging x-ray respectively. However, the phase is un detectable and unavailable for calculating the electron distribution ρ(x,y,z). The FOURIER transform of the crystal's electron distribution ρ(x,y,z) is the structure factor F(hkl) and vice versa. The structure factor can be expressed as a function of hkl:

$$F\ (hkl) = V \iiint_0^1 \rho\ (xyz)\ e^{2\pi i\ (hx+ky+lz)}\, dx\ dy\ dz \qquad \text{(equation - 4)}$$

According to the BRAGG equation, diffraction occurs only in discrete directions. Hence for calculating the electron distribution function, a summation of individual structure factors can be used to replace the integration:

$$\rho\ (xyz)\ = \frac{1}{V} \sum_h \sum_k\ \sum_l\ |F\ (hkl)|\ e^{-2\pi i\ (hx+ky+lz)} \qquad \text{(equation - 5)}$$

The structure factor being a vectorial quantity can be expressed as the complex number:

$$F_{hkl}\ = |F_{hkl}|\ e^{i\alpha_{hkl}} \qquad \text{(equation - 6)}$$

Where α is its phase and $|F_{hkl}|$ is the magnitude of the structure factor. The direct determination of structure factors and thereby the electron density becomes impossible due the lack of phase information in the crystallographic data. However, the amplitude of the structure factor can be determined from the intensity as they are proportional to each other (equation 7).

$$I_{hkl} \sim |F_{hkl}|^2 \qquad \text{(equation - 7)}$$

In this study, molecular replacement was used to solve the phase problem. Molecular replacement is a crystallographic approach of phase determination by using the already solved structure of a structurally homologous protein which serves as the starting point of refinement. The structures of two homologous proteins are likely to be similar and they show a certain degree of sequence identity. The problem of molecular replacement is to transfer and orientation of the known protein structure (also known as search model) in the new unit cell. These two molecules can be superimposed onto each other by applying rotational and translational movements. The translation is usually described by a set of distance coordinates whereas the rotation is described by a set of EULERian or polar angles.

Finding these two sets of coordinates presents a six dimensional problem which can be split into two independent searches, first the correct rotation search followed by the search for correct translation. To do so, a PATTERSON map is calculated from both the search model and the experimental data. The PATTERSON map is a vector map representing inter-atomic distances as vectors from an origin. From the experimental data it can be calculated without the knowledge of the phases as:

$$P\ (uvw) = \frac{1}{V} \sum_{hkl}\ |F\ (hkl)|^2 \cos\ [2\pi\ (hu +\ kv + lw)\rfloor \qquad \text{(equation - 8)}$$

The coefficients in equation 8, in contrast to the coefficients in equation 5, are only structure factor amplitudes and not structure factors. A rotation of the PATTERSON map results in the same rotation of the real structure and vice versa. A rotation of a PATTERSON map of the search model (resulting P_r ($\mathbf{u}_r$)) and an overlap function of the PATTERSON maps from the search model and the data set is calculated as:

$$R\ (\alpha\beta\gamma) =\ \int_u P\ (\boldsymbol{u}) \times P_r\ (\boldsymbol{u}_r) d\boldsymbol{u} \qquad \text{(equation - 9)}$$

Where U is the volume of the Patterson map where the self-Patterson peaks (i.e. the peaks from intra-molecular vectors) are located. The product function R will be maximum for the correct orientation and is dependent on the rotation angles α, β, γ. The rotation function can be determined by conducting a search for the maximum overlap of the two Patterson maps of the search model and the unknown structure. Rossmann and Blow were the first to present the mathematical algorithm in 1962 to solve the rotation function. However, nowadays the most commonly used rotation function is the one introduced by Cowther in 1972 which is much faster.

After finding the correct rotation function, the correct position of the search model in the asymmetric unit needs to be determined. This is also done in the Patterson space by maximizing R of equation 9 but this time using inter-molecular vectors. In this case, R is dependent on translational vectors and not on angles. If the search model only partially forms the contents of the asymmetric unit, - e.g. if the search model exists as a monomer whereas the protein being studied is a dimer – then the rotation function is expected to have several solutions. A translation function needs to be calculated for each solution of the rotation function. In this investigation, MOLREP (Murshudov *et al.*, 1997) was used to carry out molecular replacement.

3.3.5. Model building and Refinement

The observed intensities and model phases were used to calculate an initial electron density (equation 5). However, due to the disagreements between the search model and the structure, this initial electron density harbored many errors and was probably affected by model bias as well. The first coordinates were checked and adjusted against $2F_o$-F_c and F_o-F_c maps in COOT (Emsley *et al.*, 2010) and improved by manually rebuilding refining in REFMAC (Murshudov *et al.*, 1997). Calculation of R-factor

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - k\, |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|} \qquad \text{(equation - 10)}$$

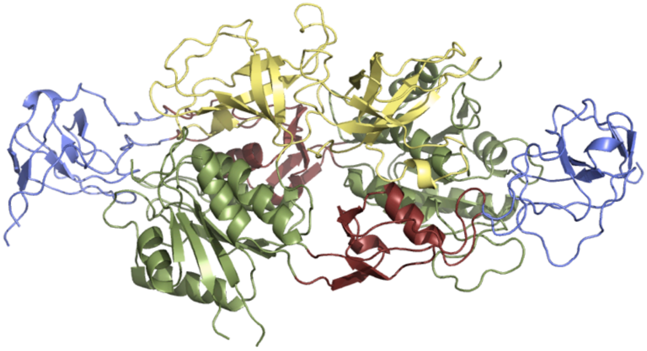
was used to monitor the agreement of observed and calculated structure factors. R_{free} was used to observe the progress of the refinement. It is calculated as the conventional R-factor but uses 5% of the observed reflections which were excluded from refinement. Refinement was continued till the R-factors converged.

3.3.6. Bioinformatics and tools:

Theoretical data about the protein constructs such as extinction coefficients were calculated using the Protparam tool (Gasteiger *et al.* 2003). ClustalW (Larkin *et al.* 2007) as incorporated in the T-coffee server (Notredame 2010) was employed to generate multiple sequence alignments and figures of these alignments were produced with ESpript (Gouet *et al.* 1999). Pymol was used to generate the figures of the protein models. All protein models were deposited in the Protein databank (PDB) (Berman *et al.* 2000).

4. Results

4A. Role of Cnx1E in the synthesis of molybdenum cofactor



4A.1. Protein production

Expression clones of Cnx1E were kindly provided by Dr. Tobias Kruse (TU-Braunschweig). The protein was expressed as C-terminal strep-tag II fusion protein in *E. coli* RK5204. The cells were grown in LB media and protein expression was induced by IPTG (Figure 14A). After cell lysis, the recombinant protein was purified by StrepTactin affinity chromatography. The protein was not subjected to any further purification since it was reasonably pure after the single step purification as seen from the SDS-PAGE analysis (Figure 14B).

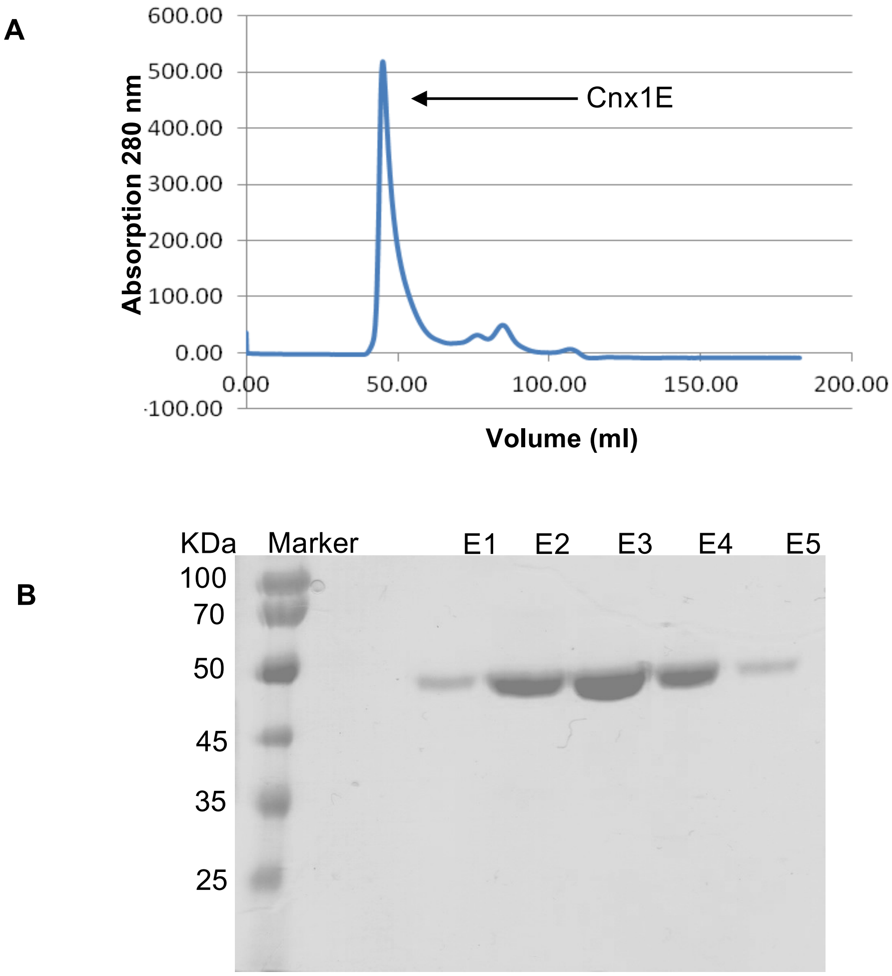


Figure 14: Purification of wildtype Cnx1E. (A) Chromatogram of StrepTactin affinity purification. (B) SDS-PAGE analysis of the peak fractions. E1-E5 are the eluted fractions corresponding to the major peak marked as Cnx1E.

4A.2. Crystallization

Crystallization of Cnx1E was first attempted with the robot by sitting-drop vapor-diffusion method under oxic conditions. The protein readily crystallized in various conditions of the JCSG core IV screen. Best crystals were seen in 50 mM Li_2SO_4, 0.1 M Hepes pH 7.5, 30% (v/v) PEG 600, 10% (v/v) Glycerol (JCSG-D11)(Figure 15A). Initial crystals from the screen appeared within 48 hours. However, these crystals were too small and showed very weak diffraction. Hence, the crystals were further optimized manually by hanging-drop vapor- diffusion method. The parameters which were mainly altered were the protein and precipitant (PEG 600) concentrations as well as the pH of the buffer (Hepes). Better crystals were observed in the hanging drop plates in 3-6 days. The protein had a tendency to form better crystals at a concentration of 4 mg/mL and at a pH 4-5 with lower precipitant concentrations. The optimization resulted in much bigger crystals (about 40µ) in the condition, 50 mM Li_2SO_4, 0.1 M Hepes pH 4.5, 15% (v/v) PEG 600, 10% (v/v) Glycerol (Figure 15B). All crystals were transferred into cryo-solution containing reservoir buffer with 16 % (v/v) glycerol as cryo-protectant and flash-cooled at 100 K in liquid nitrogen.

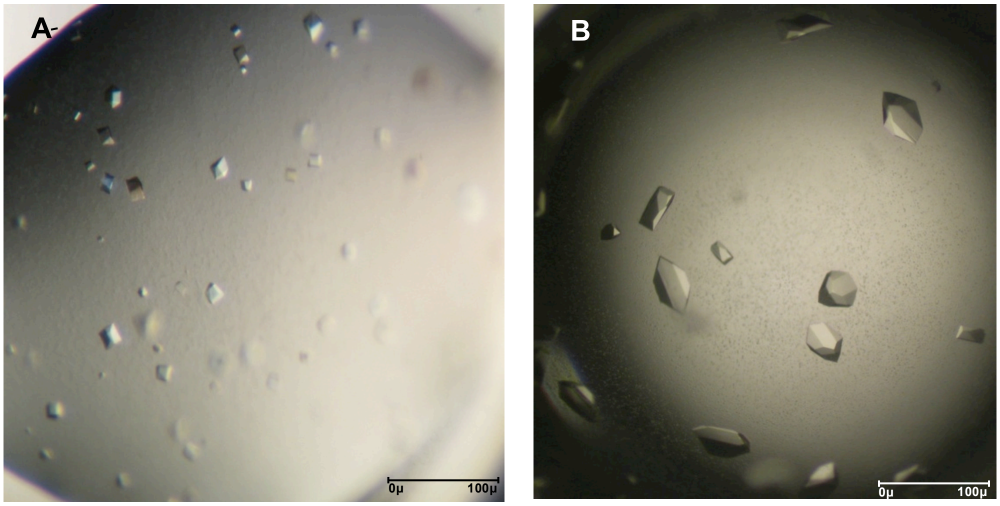


Figure 15: Crystals of Cnx1E. (A) Initial crystals of Cnx1E obtained in the D11 condition of the JCSG core IV screen in the 96 well sitting-drop set-up. (B) Manually optimized crystals of Cnx1E obtained in the 24 well hanging-drop set-up.

4A.3. Data Collection, phasing, refinement and validation

The dataset for Cnx1E crystal was collected in house at a Rigaku MicroMax 007HF rotating anode X-ray generator. The crystal belonged to the space group I222 and datasets up to 2.6 Å resolution were collected. The datasets were processed with XDS (Kabsch, 2010) and scaled with SCALA (Evans, 2006) as included in the CCP4 suite (CCP4, 1994). The Mathews coefficient (Mathews, 1968) was determined to V_M = 2.8 $Å^3$ Da^{-1}, which indicated the presence of only one Cnx1E molecule in the asymmetric unit. MOLREP (Murshudov *et al*., 1997) was used to perform molecular replacement using the monomer of MoeA (PDB ID-1FC5) (Schrag *et al*., 2001) as the search model. The solution obtained by MOLREP was manually corrected and rebuilt in COOT (Emsley *et al*., 2010). Following this, the model was refined using REFMAC5 through rigid body and restrained refinement (Murshudov *et al*., 1997). After several rounds of refinement R_{work} and R_{free} finally converged to 26.6% and 29.3%, respectively. Table 1 summarizes the data collection and refinement statistics. Validation of the final models was done using MOLPROBITY (Davis *et al*., 2007). All the residues for every model fall into the allowed region of the Ramachandran plot (Ramachandran & Sasisekharan, 1968).

Table 1: Crystallographic data and refinement statistics

.

Data collection	
Beamline	Rigaku MircoMax 007HF
Wavelength (Å)	1.541
Resolution (Å)[a]	30.92 -- 2.6 (2.693-2.6)
Space group	I222
Unit cell parameters	
a,b,c (Å)	66.18, 124.03, 131.07
α,β,γ (°)	90, 90, 90
Unique reflections[a]	16,123 (1559)
Multiplicity[a]	3.8 (3.9)
Completeness (%)[a]	98.4 (97.2)
I/σI[a]	20.18(2.57)
R_{merge}[a,b]	0.075 (0.504)
Wilson B (Å^2)	41.04
Solvent content (%)	46.4
Refinement	
R_{work}[c]	0.2664
R_{free}[d]	0.2931
No. of atoms	
Protein	3115
Water	131
Total	3246
r.m.s.d from ideal	
Bond lengths (Å)	0.013
Bond angle (°)	1.67
Ramachandran plot	
Favored (%)	96.1
Allowed (%)	2.2
Disallowed (%)	0.0

[a] Values in parentheses account for the highest resolution shell.

[b] $R_{merge} = \Sigma|Io - \langle I \rangle| / \Sigma Io$

[c] $R_{work} = (\Sigma||Fo|-|Fc||/\Sigma|Fo|)$; |Fo|: structure factors magnitudes observed; |Fc|: structure factors calculated.

[d] R_{free} calculates as R_{work} but from 5% randomly selected reflections that are excluded from refinement

4A.4. Overall structure of Cnx1E monomer

The Cnx1E monomer composed of 460 amino acid residues, adopts an extended 'L' shape consisting of four well separated sub-domains (Figure 16A, B). The long arm of the letter 'L' comprises of sub-domains 1 and 2, while sub-domain 4 forms the base. Sub-domain 3 is at the corner of the two arms. The overall structure of the monomer is shown in Figure 16A. From the crystal structure of Cnx1E it is seen that sub-domain 3 is the largest of the four whereas the smallest of them is sub-domain 1 which is also involved in bridging the sub-domains 2 and 3. The extended ‘L’ shape of the monomer also aids in its dimerization. This is discussed later.

From the primary structure of the protein it is seen that sub-domains 1 and 3 are formed of discontiguous amino acid residues. The N-terminal end of the polypeptide chain folds into an α-helix, forming a part of the sub-domain 3, and continues further to form half of sub-domain 1 and the complete sub-domain 2. After folding into sub-domain 2, the chain extends to form the remaining half of sub-domain 1, followed by the major portion of sub-domain 3 and concludes in sub-domain 4. As a result of this discontiguity, there are two linkers involved in joining sub-domains 1 and 2, as well as sub-domains 1 and 3, whereas sub-domain 4 is connected to sub-domain 3 by a single linker.

The overall structure of Cnx1E shows that the number of β-strands outnumbers the number of α-helices by a ratio of 1:2 and almost one third of the molecule is comprised of flexible loops (Figure 16A). While sub-domain 3 has almost an equal number of helices as β-strands, the remaining sub-domains predominantly consist of β-strands. In the figures of the protein molecule, the helices are represented by letters whereas the β-strands are denoted by numbers. The same form of representation is used for all the structures in this work. The structure of each of the four sub-domains is described below in detail.

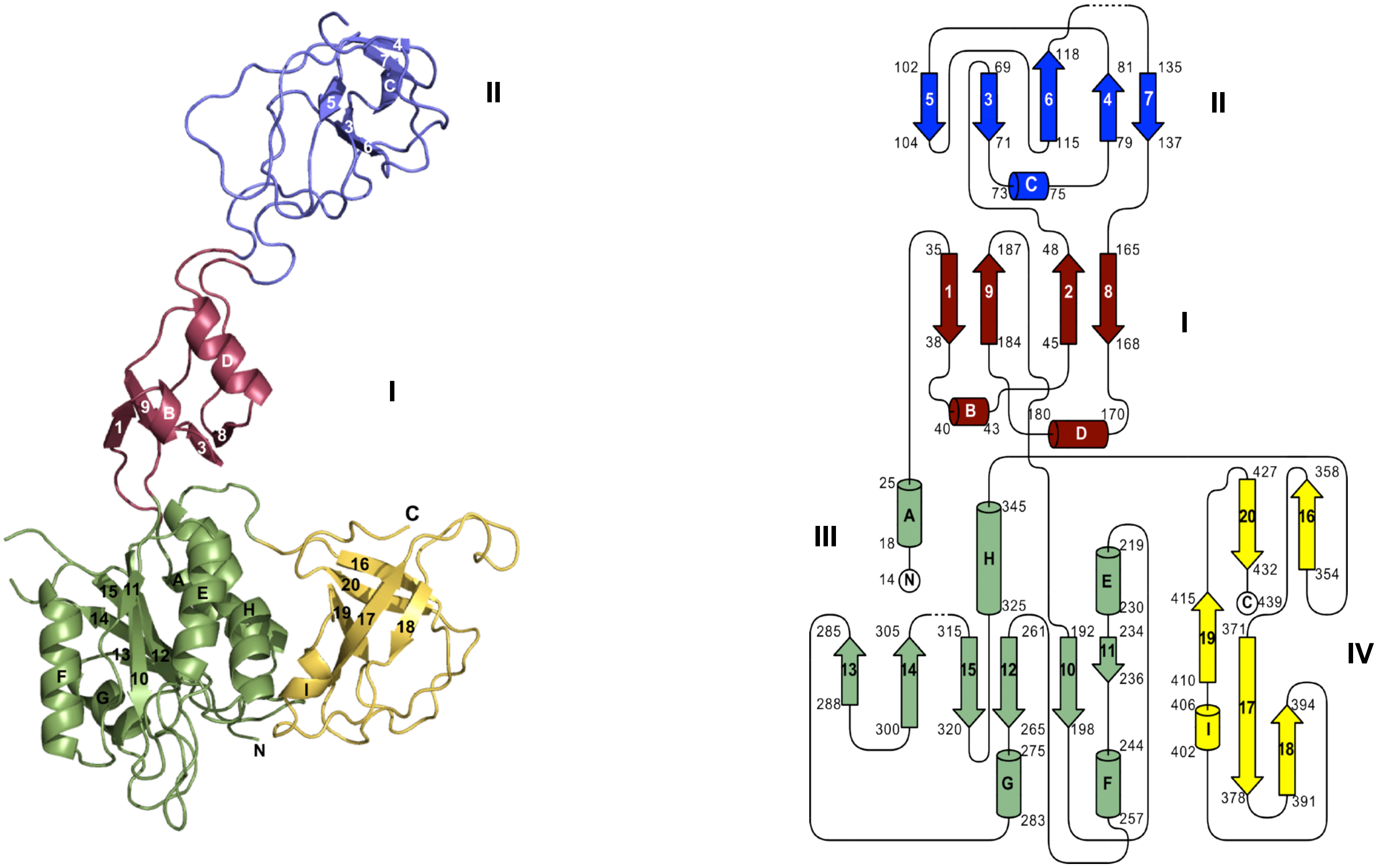


Figure 16: Structure of Cnx1E. (A) Monomer of Cnx1E (content of the asymmetric unit) in cartoon representation. The four separate sub-domains are differentiated by different colors. (Sub-domain 1– red, sub-domain 2- blue, sub-domain 3-green and sub-domain4- yellow) The same color code has been used to represent these domains in all the structures of Cnx1E depicted in this work. The helices are represented by letters whereas the β-strands are denoted by numbers. (B) Topology diagram of Cnx1E. Helices are depicted as cylinders and β-strands as arrows. The individual domains are colored according to figure 16A.

Sub-domain 1 (Figure 17A), the smallest of the sub-domains, consists of two polypeptide segments, composed of 20 and 29 amino acid residues (35-55 and 159-188) running anti-parallel to each other. In the amino acid sequence, these two segments are separated from each other by 85 residues that make up sub-domain 2. The sub-domain is composed of four β-strands and two α-helices. This kind of two-stranded β-structure with hydrogen bonding pattern similar to β-sheets are generally referred to as β-ladders. From the overall quaternary structure of this domain, it appears that the two polypeptide chains fold into a loop consisting of a right handed spiral. This spiral starts with a β-ladder composed of β-1 and β-9, followed by two short α-helices, B and D, and ends with the second β-ladder of β-2 and β-8 extending towards sub-domain 2.

The edge of the long arm of the ‘L’ shaped molecule is occupied by sub-domain 2 (Figure 17B). Unlike sub-domain 1, it is formed of contiguous amino acid residues (residues 58-153). This segment is mainly made up of five β-strands. Two of them form a β-ladder (β-4 and β-7) and the remaining 3 strands form a central mixed β-sheet (β-3, β-5 and β-6). Another characteristic feature of this sub-domain is the presence of a considerable amount of loops.

Sub-domain 3 (Figure 17C), the largest of the four sub-domains, has a compact globular α/β/α structure. Residues 14-32 from the N-terminal end, which forms an α-helix, pack with residues 189-349 to complete this sub-domain. It is characterized by a central mixed β-sheet composed of 6 β-strands. The two faces of this β-sheet are occupied by four α-helices, two on each side. The four α-helices are formed from the contiguous segment of the amino acid sequence and pack almost parallel to the central β-strands. The α-helix A at the N-terminal end of the monomer, predominantly consists of polar amino acids and is exposed to the solvent. It is located alongside the α-helices E and H. It was further observed that the structure of sub-domain 3 is highly similar to Cnx1G. This is discussed in detail in the next section.

The short arm of the ‘L’ shaped monomer is formed by sub-domain 4 (Figure 17D) and is located at the C-terminal end of the molecule. This domain is predominantly composed of five β-strands along with a 3_{10} helix. The characteristic feature of this domain is the β-barrel formed by twisting of the five β-strands. The longest among these five strands is the strand β-17 whose two ends (N and C terminus) interact with two different β-strands (β-18 and β-19). Its C-terminal end is involved in the formation of a β-hairpin along with the strand β-18 while its N-terminal end packs along the β-strand β-19. The strands β-18 and β-19 are connected to each other by a long loop that harbors a short and the only α-helix I of the domain. Two loops, one between β-16 and β-17 and the other between β-19 and β-20, running in anti-parallel direction form a cap at one end of this β-barrel. This extended conformation of two loops is reminiscent of antiparallel β-strands, however they lack extensive hydrogen bonds.

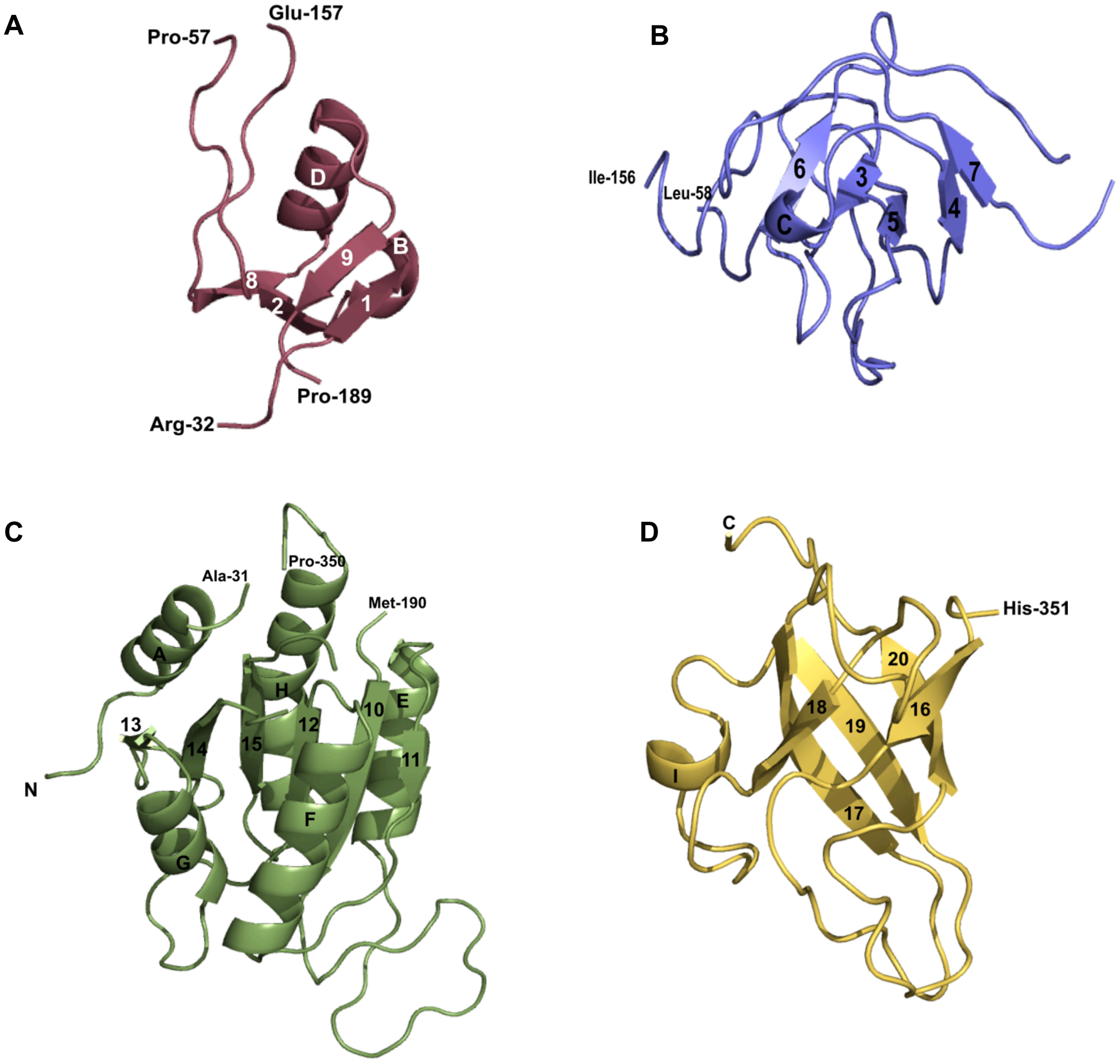


Figure 17: Cartoon representation of the four individual sub-domains of Cnx1E. (A) Sub-domain 1, (B) sub-domain 2, (C) sub-domain 3 and (D) sub-domain 4. The helices and β-strands are numbered according to Fig 12 A,B.

The crystal structure of Cnx1E revealed that the molecule exists as a homodimer (Figure 18). The extended shape of the molecule facilitates all four sub-domains of the monomer to indulge in extensive contacts with the other monomer for the dimerization. The monomers mainly exhibit van der Waals contacts. On closer observation of the molecular surface, numerous grooves, clefts and depressions are seen that aid in the interactions between the two monomers (Figure 18). From the crystal structure, it is seen that the sub-domains 1 and 2 of one monomer extend along the groove formed at the interface between the sub-domains 3 and 4 of the other monomer. Sub-domain 2, compared to the other three sub-domains, demonstrates least participation in intermolecular interaction and flanks from the two ends of the dimer.

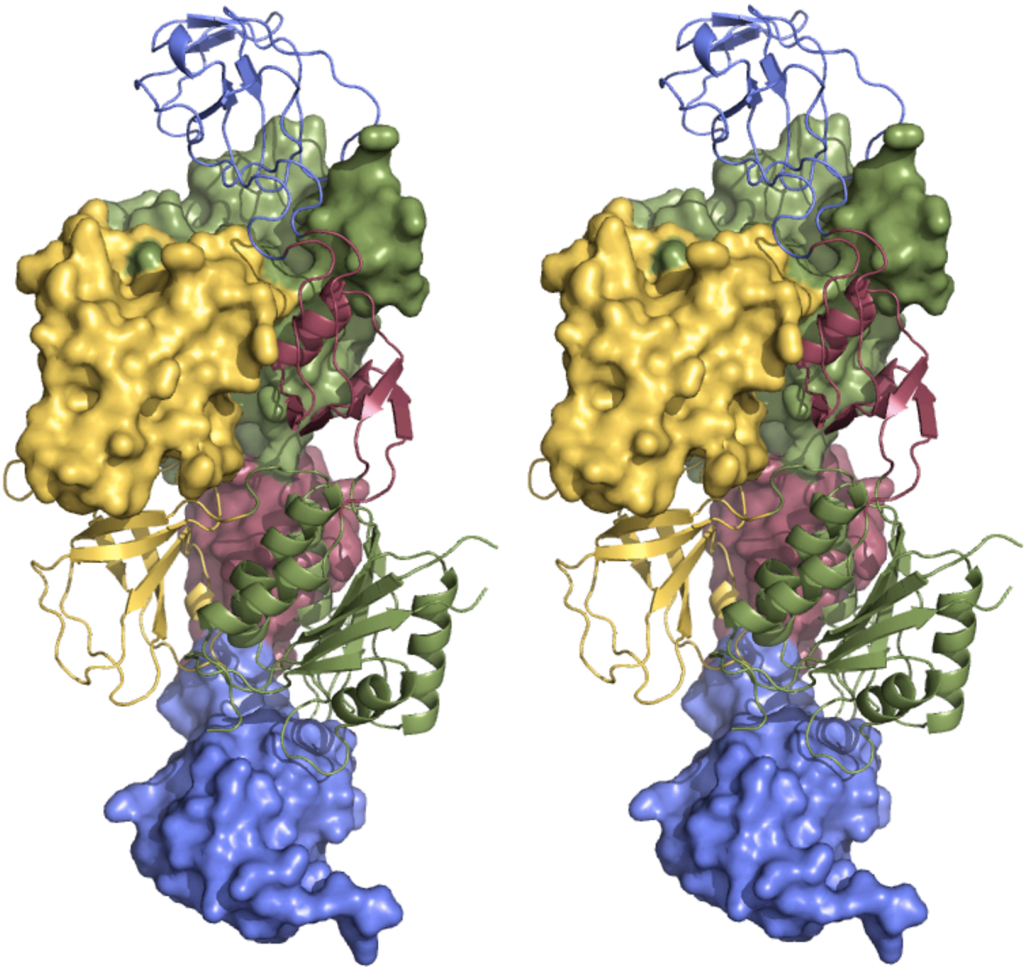


Fig 18: Stereo view of the homodimeric assembly of Cnx1E. One of the monomers of the dimeric molecule is shown in cartoon representation whereas for the other monomer, its surface is depicted.

4A.5. Structural comparison of Cnx1G with the domain 3 of Cnx1E

The structure of Cnx1G monomer is highly similar to the sub-domain 3 of Cnx1E, as indicated by root mean square distance (rmsd) of C_α positions of 1.16 Å, calculated with PYMOL. Both are compact globular structures exhibiting α/β/α architecture and are characterized by a central mixed β-sheet surrounded by α-helices. On comparing the topology (Figure 19A, B) of the two structures, it is observed that the central β-sheet of Cnx1E sub-domain 3 is composed of six β-strands, five of which are parallel and one is anti parallel, whereas in Cnx1G the central β-sheet is five stranded, of which four strands are parallel and one strand is anti-parallel. The β-strand missing in Cnx1G is β-13. In both the structures, the α-helices pack almost parallel on either side of the central β-sheet. In Cnx1E, two α-helices occupy one face of the central β-sheet, whereas the other face is occupied by three helices. On comparing the arrangement of helices in Cnx1G with the structure of sub-domain 3 of Cnx1E, it is seen that the latter has one helix less. The extra helix in Cnx1G is a 3_{10} helix (*X*), and is the only helix in the structure that is not parallel to the central β-sheet.

The presence of one extra β-strand and the absence of one α-helix in Cnx1E become clearer on superimposing Cnx1G on the sub-domain 3 of Cnx1E. Figures 19C and D show two opposite face views of the superimposition of Cnx1G on the sub-domain 3 of Cnx1E. The region of the polypeptide chain forming the extra β-strand, β-13 and the loop joining β-13 to β-14 in Cnx1E, folds into a long α-helix (*Y)* in Cnx1G. Another difference between the two structures is observed in respect to the β-strand, β-11 in Cnx1E, which is much shorter than its counterpart in Cnx1G. In the superimposed structure, it is seen that most of the β-strands and α-helices of Cnx1G coincide with their structural counterparts in sub-domain 3 of Cnx1E. Only the bent α-helix (*Y)* and the 3_{10} helix (*X)* of Cnx1G do not have a structural equivalent. In Cnx1E, the additional β-strand, β-13, and the α-helix, A, do not have a structural equivalent in Cnx1G.

A sequence alignment of Cnx1G with sub-domain 3 of Cnx1E was done using CLUSTALW (Larkin *et al.,* 2007) and it revealed the identical amino acids in the two structures (Figure 20). The figure shows a sequence based alignment of secondary structure elements where it is clearly seen that the segment of the amino acid sequence involved in forming the extra β-strand, β-13 in Cnx1E, forms the α-helix *Y* in Cnx1G. The alignment also shows a significant size difference of the β-strand β-11 in both structures .

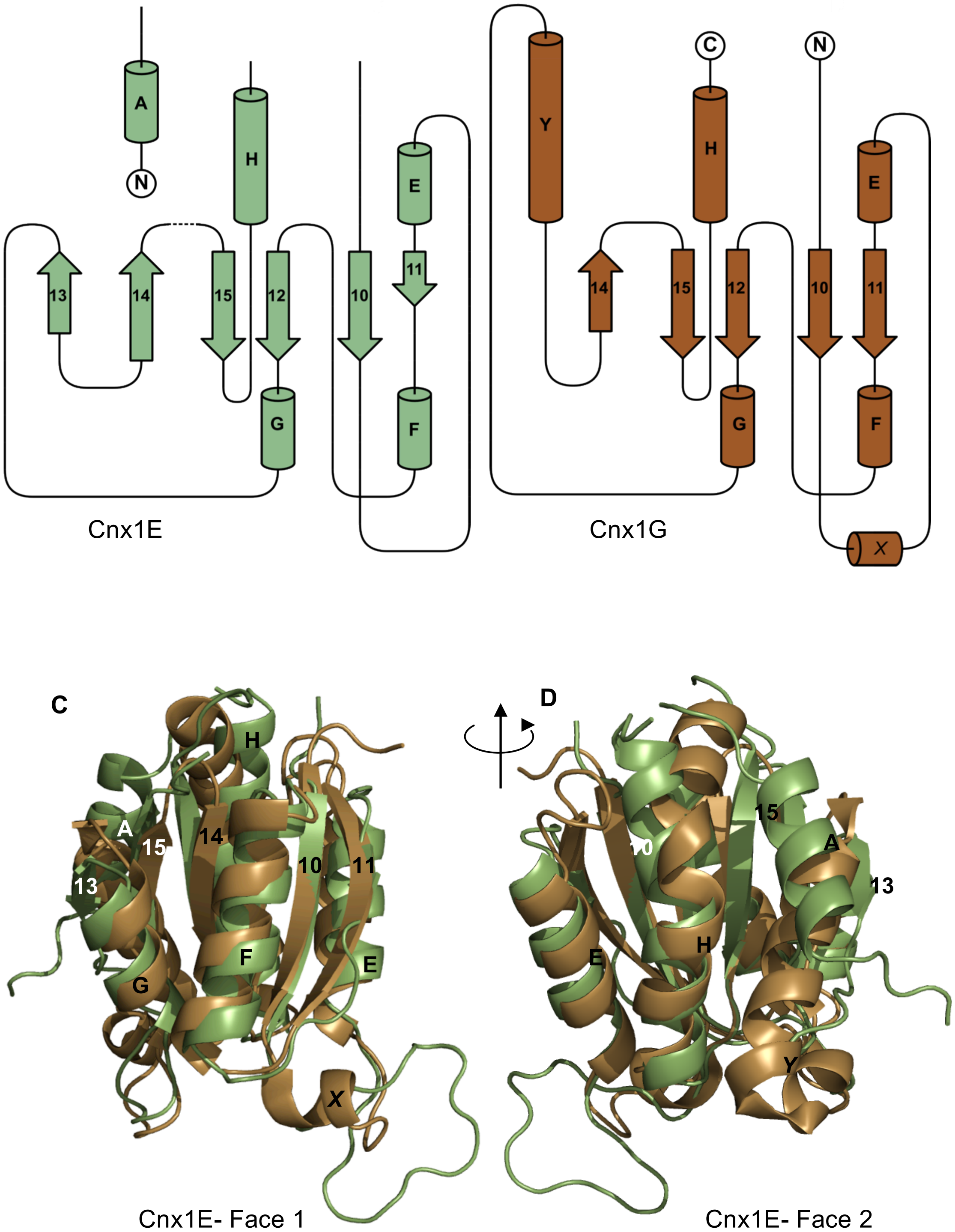


Figure19: Comparison of Cnx1E sub-domain 3 (green) with Cnx1G (brown). (A, B) Topology diagrams of Cnx1E sub-domain 3 (A) and Cnx1G (B). (C, D) Two opposite faces of superimposed Cnx1G (brown) on the domain 3 of Cnx1E in cartoon representation.

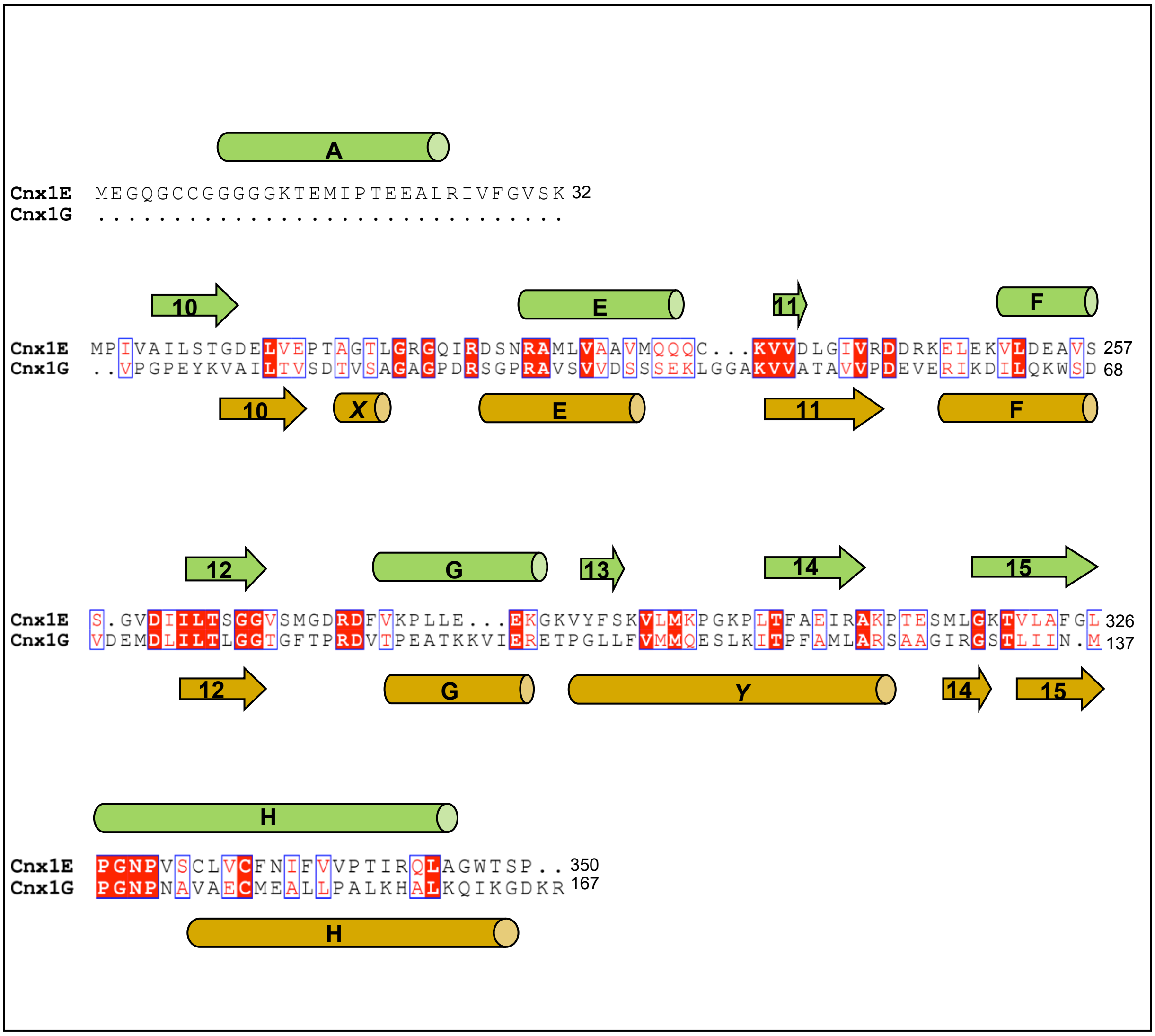


Figure 20: Structure based sequence alignment of sub-domain 3 of Cnx1E with Cnx1G. The secondary structure elements of Cnx1E are shown in green whereas for Cnx1G are shown in brown. Helices are depicted by cylinders and the β-strands are shown by arrows. The identical amino acids in the two sequence are highlighted by a red box.

4A.6. Functional characterization of residues implicated in catalysis and substrate binding

Previous studies had identified the residue Asp-274 as an important residue involved in catalysis and Gly-108 was implicated in substrate binding (Corinna Probst, *Master's Thesis*, 2012). With the aim of functional characterization of these two residues, four mutants were designed: Cnx1E-S269DD274S, Cnx1E-D274R, Cnx1E-G108DS269DD274S and Cnx1E- G108DD274R. These four mutants had the common goal of trapping the substrate MPT-AMP in the active site due to their compromised catalytic activity. Hence, from now on, they are referred as putative MPT-AMP hydrolysis mutants. The exchange of the aspartate and serine residues in the double mutant Cnx1E-S269DD274S was designed with the objective of knocking out the function of the catalytically important aspartate (Asp-274), and at the same time, making it less conservative than the Cnx1E-D274R mutant. Previously, it had been observed that the Cnx1E-G108D mutation resulted in the high saturation of the substrate in the enzyme (Corinna Probst, *Master's Thesis*, 2012). Hence, this mutation was introduced in the previous two mutants with the aim of increasing the substrate saturation in the active site of the enzyme. The mutants were expressed as C-terminal Strep-tag II fusion protein in *E. coli* SE1581 and purified by StrepTactin affinity chromatography. Figure 21 shows the SDS-PAGE analysis of the four mutants. The mutants were created by Dr. Tobias Kruse.

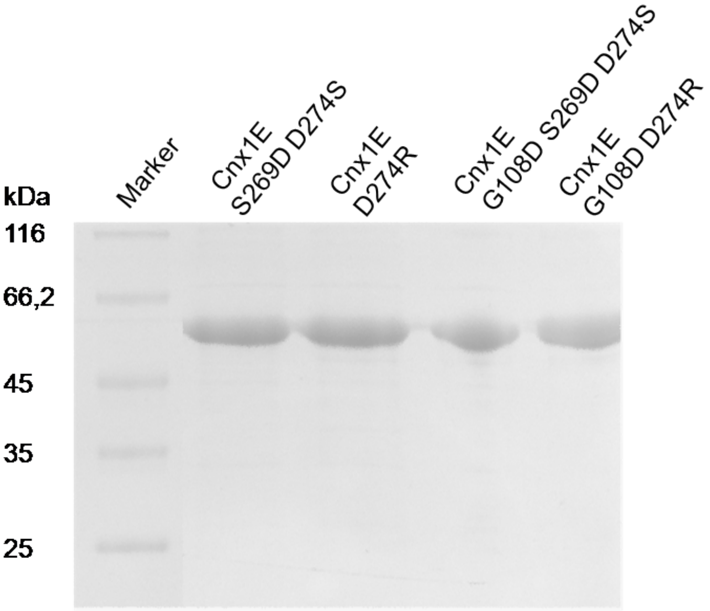


Figure 21: SDS-PAGE analysis of the four Cnx1E variants. (Figure taken from the master's thesis Corinna Probst, 2012).

4A.6.1. Analysis of MPT-AMP saturation in the putative MPT-AMP hydrolysis mutants

The mutants were analyzed for the presence of bound MPT-AMP and MPT/Moco co-purified from *E. coli.* The quantification of Moco/MPT-AMP can be performed only on its stable oxidation product FormA. Figure 22 shows the calculated peak areas of FormA per pmol of Cnx1E mutants determined by an established HPLC analysis (Corinna Probst, *Master's Thesis*, 2012). This experiment was performed by Corinna Probst as a part of her master's thesis.

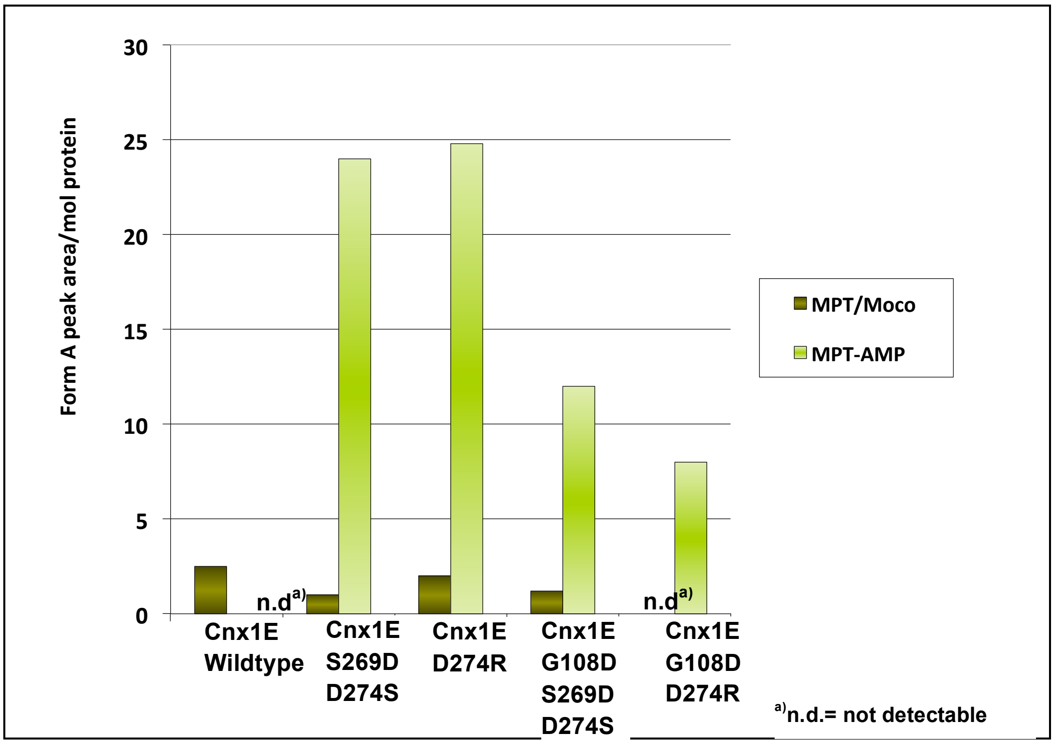


Figure 22: MPT/Moco and MPT-AMP content of the putative MPT-AMP hydrolysis mutants and the Cnx1E wild-type protein. The histograms represent MPT/Moco and MPT-AMP content in FormA peak area per pmol of protein Cnx1E-S269DD274S, Cnx1E-D274R, Cnx1E-G108DS269DD274S, Cnx1E-G108DD274R. The standard deviations were calculated from the results of three independent FormA analyzes. (Figure modified from Corinna Probst, *Master's Thesis*, 2012)

The Cnx1E wild-type protein was used as a reference. From the Figure 22, it is seen that the mutants Cnx1E-S269DD274S and Cnx1E-D274R, have comparable amounts of MPT-AMP bound to them. It is further observed that, on introducing the additional G108D mutation to Cnx1E-S269DD274S and Cnx1E-D274R, have only half as much MPT-AMP was bound to them. Very small amounts of the product MPT/Moco was observed to be present in the mutants when compared to the wildtype protein with Cnx1E-G108DD274R showing no detectable amounts of MPT-AMP by HPLC analysis. The expression, purification and biochemical analysis of bound MPT-AMP and MPT/Moco was done by Corinna Probst as a part of her master's thesis.

4A.6.2. Colorimetric detection of protein bound to the putative molybdate Cnx1E MPT-AMP hydrolysis mutants and ICP-MS analysis

Molybdate is the source of molybdenum for insertion into Moco. Previous research has shown that that MPT-AMP and molybdate bind cooperatively to Cnx1E (Llamas *et al*., 2007). Based on this observation, the determination of molybdate saturation in Cnx1E can be directly related to the MPT-AMP saturation of the enzyme. The molybdate saturation in Cnx1E was detected using a colorimetric detection method by Corinna Probst as a part of master's thesis. Figure 23 shows the Molybdate saturation in the four putative MPT-AMP hydrolysis mutants.

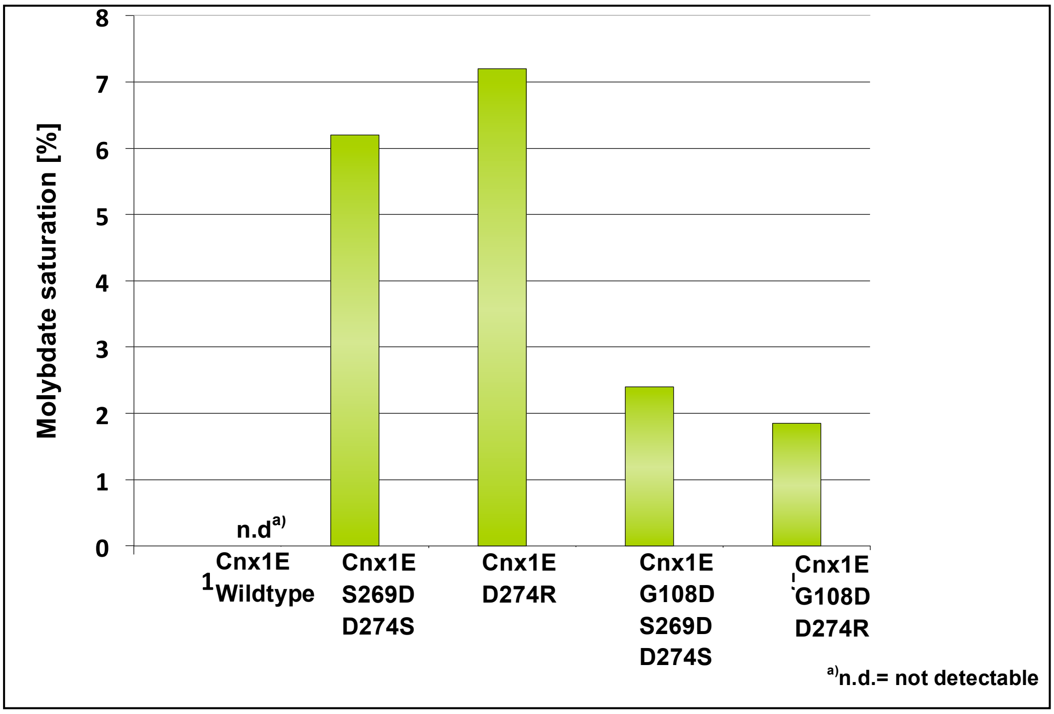
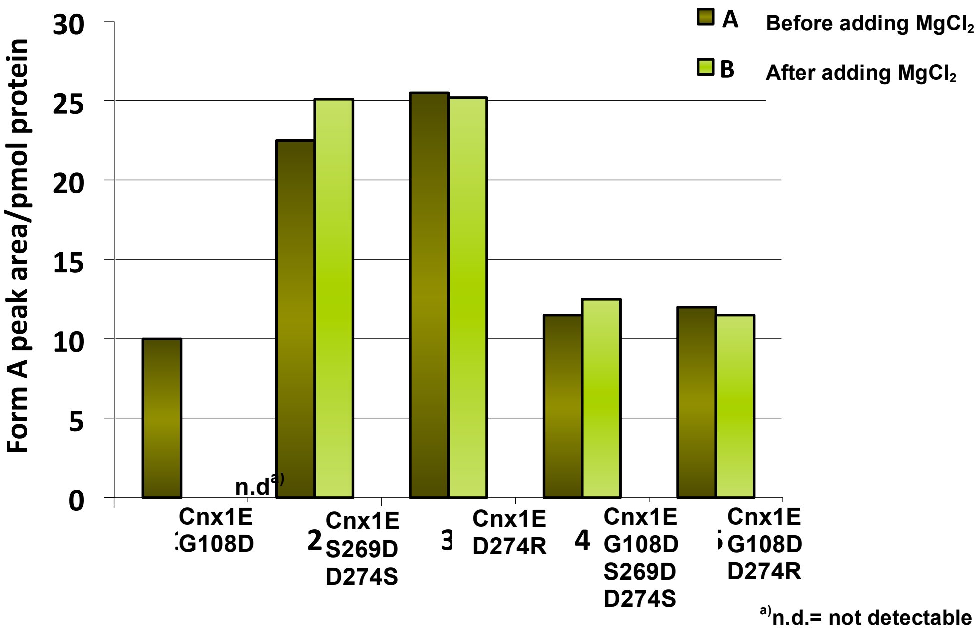


Figure 23: The Molybdate saturation in the four putative Cnx1E mutants. (Figure modified from Corinna Probst, *Master's Thesis*, 2012)

Molybdate was detected in all the four mutants. The highest molybdate saturation of about 7% was observed in Cnx1E-D274R mutant, followed by Cnx1E-S269DD274S showing about 6% saturation. The mutants with the additional G108D mutation, Cnx1E-G108DS269DD274S and Cnx1E-G108DD274R, demonstrated a low molybdate saturation of about 2%.

4A.6.3. *In vitro* MPT-AMP hydrolysis

It has already been established that Cnx1E can catalyze the hydrolysis of MPT-AMP in the presence of divalent cations followed by insertion of molybdenum to yield Moco (Llamas *et al.*, 2007). A HPLC based *in vitro* MPT-AMP hydrolysis assay was designed to determine whether the MPT-AMP hydrolysis mutants' catalytic activity was impaired. The *in vitro* MPT-AMP hydrolysis assay was performed by Corinna Probst as a part of her master's thesis. Figure 24 shows the MPT-AMP content of the four putative MPT-AMP hydrolysis and mutants before and after $MgCl_2$ incubation. As a positive control, Cnx1E-G108D co-purified with MPT-AMP was used.

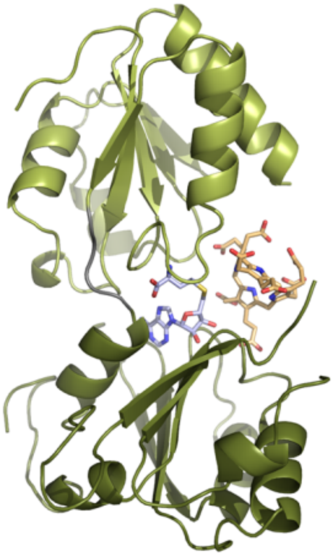


[a] n.d.= not detectable

Figure 24: ***In vitro* MPT-AMP hydrolysis assay of the putative MPT-AMP hydrolysis mutants.** The figure shows the MPT-AMP (FormA peak area per pmol protein) of the four putative MPT-AMP hydrolysis mutants along with the mutant Cnx1E G108D as the positive control before and after $MgCl_2$ incubation. (Figure modified from Corinna Probst, *Master's Thesis*, 2012)

The comparison of the detectable MPT-AMP strength (Figure 24) before and after addition of $MgCl_2$ shows that the MPT-AMP composition in the mutants is not changed at all in the four putative MPT-AMP hydrolysis mutants. However, it was observed that in the case of Cnx1E-G108D mutant which was used a positive control, no MPT-AMP could be detected after 30 minutes. Thus the four mutants cannot process MPT-AMP to Moco efficiently, which confirmed the role of Asp-274 in catalysis and the role of Gly-108 in substrate coordination.

4B. Role of NirE in the synthesis of the tetrapyrrole heme d_1



4B.1. Crystallization

Crystallization trials of purified recombinant NirE from *Pseudomonas Aeruginosa* were attempted in the presence of SAM that had been added to the crystallization set-up. SAM had a conducive effect on the growth of NirE crystals as crystallization attempts without SAM resulted in the growth of poor quality crystals. Although SAM was added to the crystallization set-up, it was converted to SAH by the enzyme. NirE readily crystallized in various conditions. The best crystals grew in Tris pH 8.0, 24 % PEG 6000 (Figure 28A) within 3-6 days. Co-crystallization of NirE along with its substrate uro'gen III and the reaction by-product SAH was also attempted by sitting-drop vapor-diffusion method. However, in this case, the crystallization was done under anoxic conditions due to the high oxygen sensitivity of uro'gen III. The reaction by-product SAH was included in the crystallization set-up instead of SAM, in order to trap uro'gen III in the active site and prevent its methylation by the enzyme. Similar to the previous crystallization trials, the NirE-SAH-uro'gen III complex crystallized in several conditions in a time span of 4-8 days. The best crystals grew in 0.2 M Li_2SO_4, 0.1 M Tris pH 8.5, 1.26 M $(NH_4)_2SO_4$(Figure 28B). The crystals of NirE-SAH, as well as those of NirE-SAH-uro'gen III were base-centered orthorhombic. Apart from a few variations in their unit cell dimensions, both the crystals were isomorphous (for details, see table 2). Crystallization of NirE-SAH and NirE-SAH-uro'gen III complex was done by Sonja Storbeck as a part of her Ph.D. thesis (Sonja Storbeck, Ph.D. Thesis, 2011).

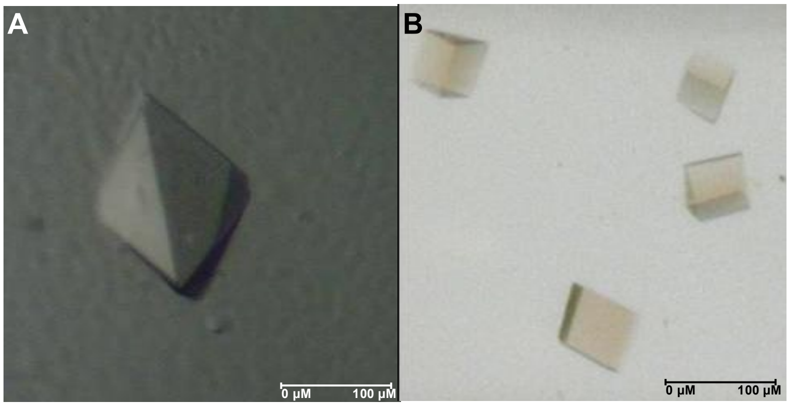


Figure 28: Crystals of NirE. (A) Crystal of NirE + SAM. (B) Crystals of NirE+SAH+Uro'gen III

4B.2. Data Collection, phasing, refinement and validation

The dataset for NirE-SAH crystal was collected in house at a Rigaku MicroMax 007HF rotating anode X-ray generator, whereas the dataset for NirE-SAH-uro'gen III crystals was collected on beamline 14.2 of the Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung. Both NirE-SAH and NirE-SAH-uro'gen III crystals belong to the space group $C222_1$ and both diffracted up to 2.0 Å resolution. The dataset of NirE-SAH was processed with HKL 2000 (Minor *et al.,* 2002) and scaled with SCALA (Evans, 2006) of the CCP4 suite (CCP4, 1994). The dataset of NirE-SAH-uro'gen III, on the other hand, was processed and scaled with XDS and XSCALE (Kabsch, 2010). The Mathews coefficient (Mathews, 1968) for the two datasets were determined as V_M = 2.29 $Å^3$ Da^{-1} and V_M = 2.44 $Å^3$ Da^{-1}, which indicated the presence of only one NirE molecule in the asymmetric unit. MOLREP (Murshudov *et al*., 1997) was used to perform molecular replacement using the monomer of CobA (PDB ID - 1S4D) (Vevodova *et al*., 2004) as the search model. The solution obtained by MOLREP was manually corrected and rebuilt in COOT (Emsley *et al*., 2010). Following this, the model was refined using REFMAC through rigid body, restrained, and TLS refinement (Murshudov *et al*., 1997). At first, several rounds of rigid body and restrained refinement were performed and the final refinement was done by TLS refinement (Painter & Merritt, 2006) using 2 and 5 TLS groups as defined by the TLS Motion Determination Server (Painter & Merritt, 2006) for NirE-SAH and NirE-SAH-uro'gen III, respectively. In case of the NirE-SAH-uro'gen III crystals, unambiguous density for uro'gen III was found. Uro’gen III was modeled into the density by generating its PDB coordinates using PRODRUG (Schuttelkopf and Aalten, 2004). After this, further refinement of the structure was performed. The R_{work} and R_{free} finally converged to 21.7% and 25.2%, respectively. The final R_{work} and R_{free} for the NirE-SAH structure converged to 23.5% and 28.7%, respectively. Validation of the final models was done using MOLPROBITY (Davis *et al*., 2007). All the residues for every model fall into the allowed region of the Ramachandran plot (Ramachandran & Sasisekharan, 1968).

Table 2: Crystallographic data and refinement statistics

	Dataset I NirE + SAH	**Dataset II** NirE + SAH + Uro'gen III
Data collection		
Beamline	Rigaku MircoMax 007HF	BESSY (BL14.2)
Wavelength (Å)	1.541	0.91841
Resolution (Å)^{a}	46.3 - 2.0 (2.05 –2.00)	35.0 - 2.0 (2.05 –2.00)
Space group	$C222_1$	$C222_1$
Unit cell parameters		
a,b,c (Å)	60.9, 115.1, 76.8	60.9, 116.2, 77.1
α, β, γ (°)	90, 90, 90	90, 90, 90
Unique reflections^{a}	19,360 (1411)	18,362 (1418)
Multiplicity^{a}	4.4 (4.5)	3.9 (4.0)
Completeness (%)^{a}	98.82 (99.86)	92.6 (99.9)
$I/\sigma I^{a}$	20.19 (2.58)	22.8 (4.0)
$R_{merge}{}^{a,b}$	0.043 (0.586)	0.083 (0.451)
Wilson B (Å^2)	35.1	34.4
Solvent content (%)	46.4	49.7
Refinement		
$R_{work}{}^{c}$	0.2349	0.2169
$R_{free}{}^{d}$	0.2868	0.2526
No. of atoms		
Protein	1,864	1,825
Ligand	26	86
Water	195	129
Total	2,085	2,040
Root mean square deviation from ideal		
Bond lengths (Å)	0.0239	0.0219
Bond angle (°)	2.167	1.935
Ramachandran plot		
Favored (%)	97.1	97.0
Allowed (%)	2.9	3.0
Disallowed (%)	0.0	0.0

[a] Values in parentheses account for the highest resolution shell.

[b] R_{merge}= Rmerge = $\Sigma|I_o - \langle I \rangle| / \Sigma I_o$

[c] R_{work}= $(\Sigma||F_o|-|F_c|| / \Sigma|F_o|)$; $|F_o|$: structure factors magnitudes observed; $|F_c|$: structure factors calculated.

[d] R_{free} is calculated as R_{work} but from 5% randomly assigned reflections that are excluded from refinement.

The principal objective of this investigation was to obtain the enzyme-substrate complex structure of NirE with uro'gen III. To achieve this, NirE was incubated with 82-fold molar excess of uro’gen III that had been prepared by chemical reduction of uroporphyrin III with H_2/palladium. In addition, the reaction by-product SAH was added to the NirE-uro’gen III mixture, instead of SAM. The objective of this substitution was to prevent the catalytic conversion of uro'gen III and thereby trap the substrate in the active site of the enzyme. The resulting crystals of NirE from these crystallization set-ups had SAH and uro’gen III bound in its active site. In the refined NirE-uro’gen III complex structure, the localization, orientation, and conformation of the substrate was clearly defined by the additional electron density. The uro’gen III molecule binds above SAH and has partial contact with the solvent (Figure 29). In the omit map, initial electron density was clearly visible for the tetrapyrrole macrocycle, as well as for the acetate and propionate substituents of two opposite pyrrole rings (Figure 29).

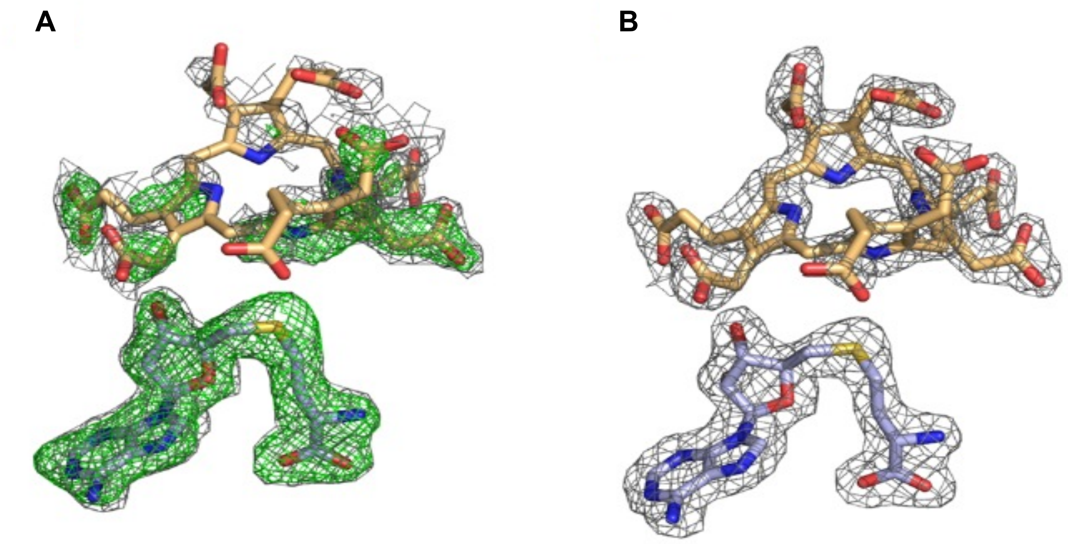


Figure 29: Substrate binding site of NirE. (A) Omit maps of SAH and uro’gen III calculated in the absence of the substrates from the final structure model. The $2F_o$-F_c map is contoured to 1 σ and represented by a black mesh. The F_o-F_c map is contoured to 3 σ and represented by a green mesh. (B) $2F_o$-F_c map of SAH and uro’gen III calculated in the presence of the substrates shown as black mesh contoured to 1 σ.

4B.3. Structure of NirE

The NirE monomer is composed of two similarly sized domains (N-terminal domain, A and C-terminal domain, B) joined by a short linker (residues 130-133) (Figure 30A, B). Both domains are characterized by a central five-stranded β-sheet surrounded by five α-helices. The β-strands of the N-terminal domain A (residues 1-129) are parallel and arranged in the order β-5 β-1 β-4 β-2 β-3 (the numbering represents the sequential appearance of β-strands in the NirE primary structure).The α-helices A, B, and C are located on the same side of the β-sheet, whereas D and E lie on the opposite side. Similar to the N-terminal domain A, the C-terminal domain B (residues 134-279) exhibits the same arrangement of secondary structure elements, with a central five-stranded β-sheet core surrounded by five α-helices. However, the five β-strands of domain B are not arranged parallel to each other. In the mixed β-sheets of domain B, the strands are arranged in the order β-6 β-7 β-10 β-8 β-9. The α-helices F and J are located on one face of the mixed β-sheet, whereas α-helices G, H, and I occupy the opposite face. Although the two domains of the monomer exhibit the same arrangement of secondary structure elements, they differ in their topology (Figure 30A, B

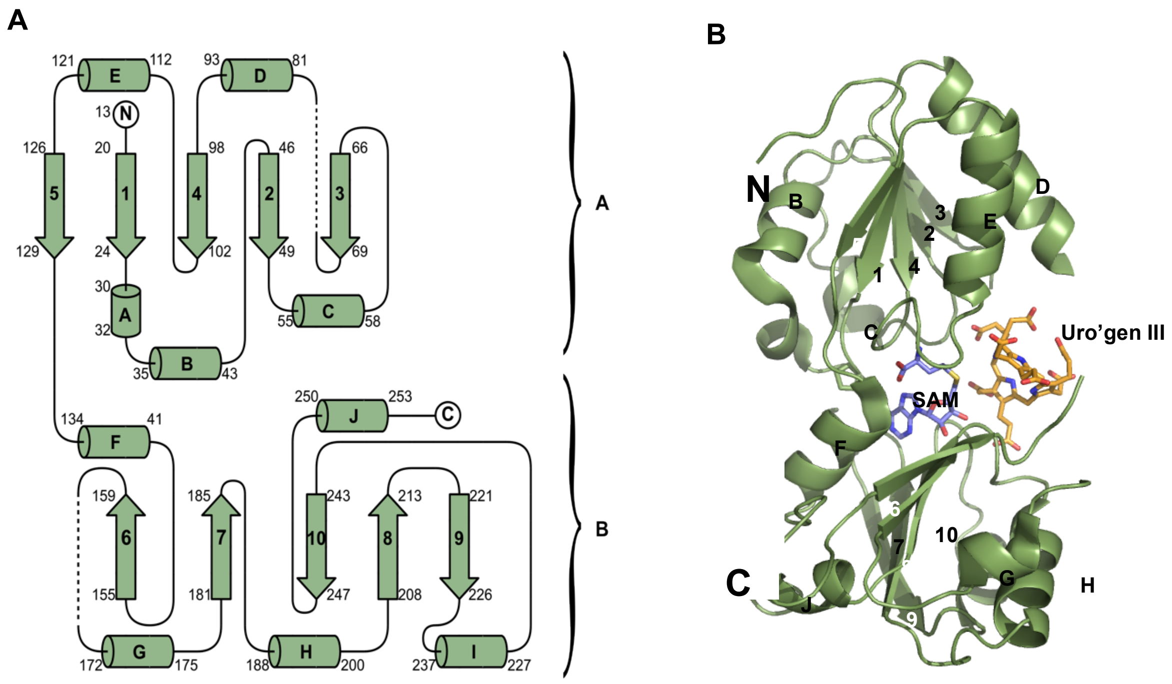


Figure 30: Structure of NirE. (A) Topology diagram of NirE. Helices are depicted as cylinders and β-strands as arrows. The N-terminal and C-terminal domains are labeled as A and B, respectively. (B) Monomer of NirE (content of the asymmetric unit) in cartoon representation with SAH (light blue) and uro'gen III (light orange) shown as sticks.

The crystal structure revealed a homodimeric organization of NirE, which is in accordance with its dimeric state in solution. In the crystal structure, the homodimeric assembly of the protein molecule is created by a two-fold crystallographic symmetry axis (Figure 31A, B). In the NirE dimer, the five stranded β-sheets of both the C-terminal domains combine to form a single, slightly twisted ten-stranded β-sheet. The dimer has a total buried surface area of 5140 Å^2. An interesting feature of this dimeric assembly is the long loop between α-helix F and β-strand, β-6 of domain B that extends into the active site pocket of the other monomer and participates in substrate binding and also plays a role in catalysis. In Figure 32A and B, this loop is highlighted in red. From the figure, an arginine residue is clearly seen to interact with uro'gen III at the active site of the other monomer. The dimeric arrangement and overall fold for NirE in the presence as well as in the absence of the bound substrate is highly similar. Apparently, there seem to be no marked conformational changes in the NirE backbone upon binding of uro'gen III.

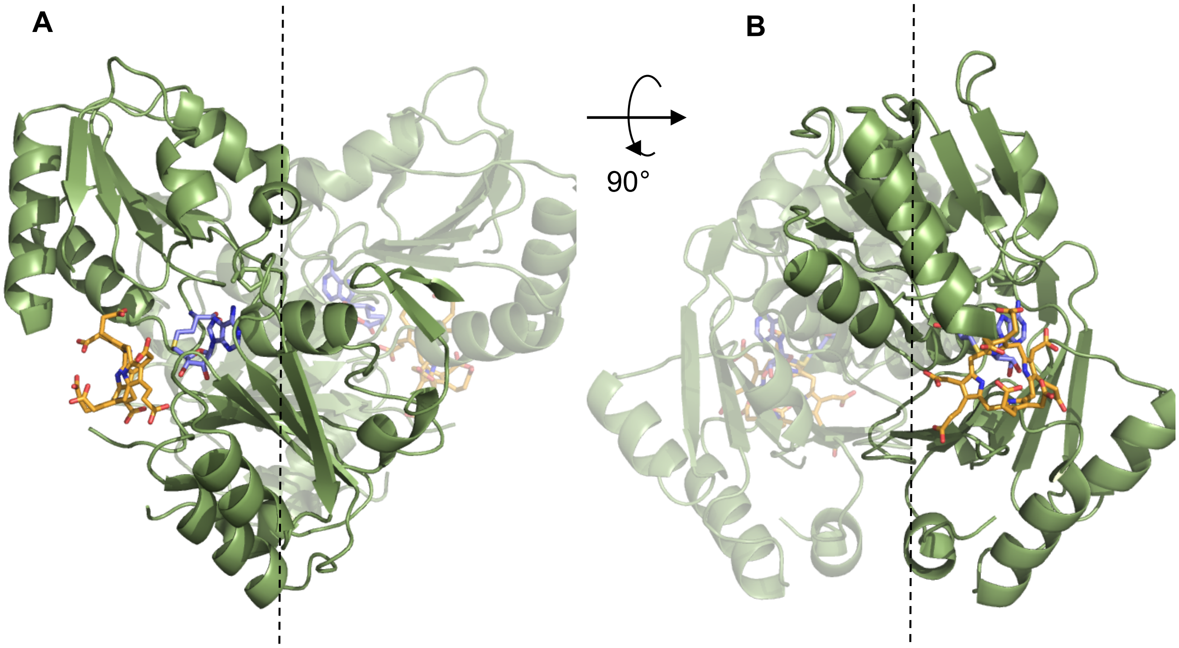


Figure 31: Dimeric organization of NirE. Cartoon representation of the physiological dimer of NirE created by a crystallographic symmetry operation (rotation axis is indicated by a dashed line). In the dimer, one of the two monomers is shown in a transparent shade of green. SAH and uro'gen III are represented as in Figure 29B. The two different views of dimeric NirE are related by a rotation about the two-fold symmetry axis of approximately 90°.

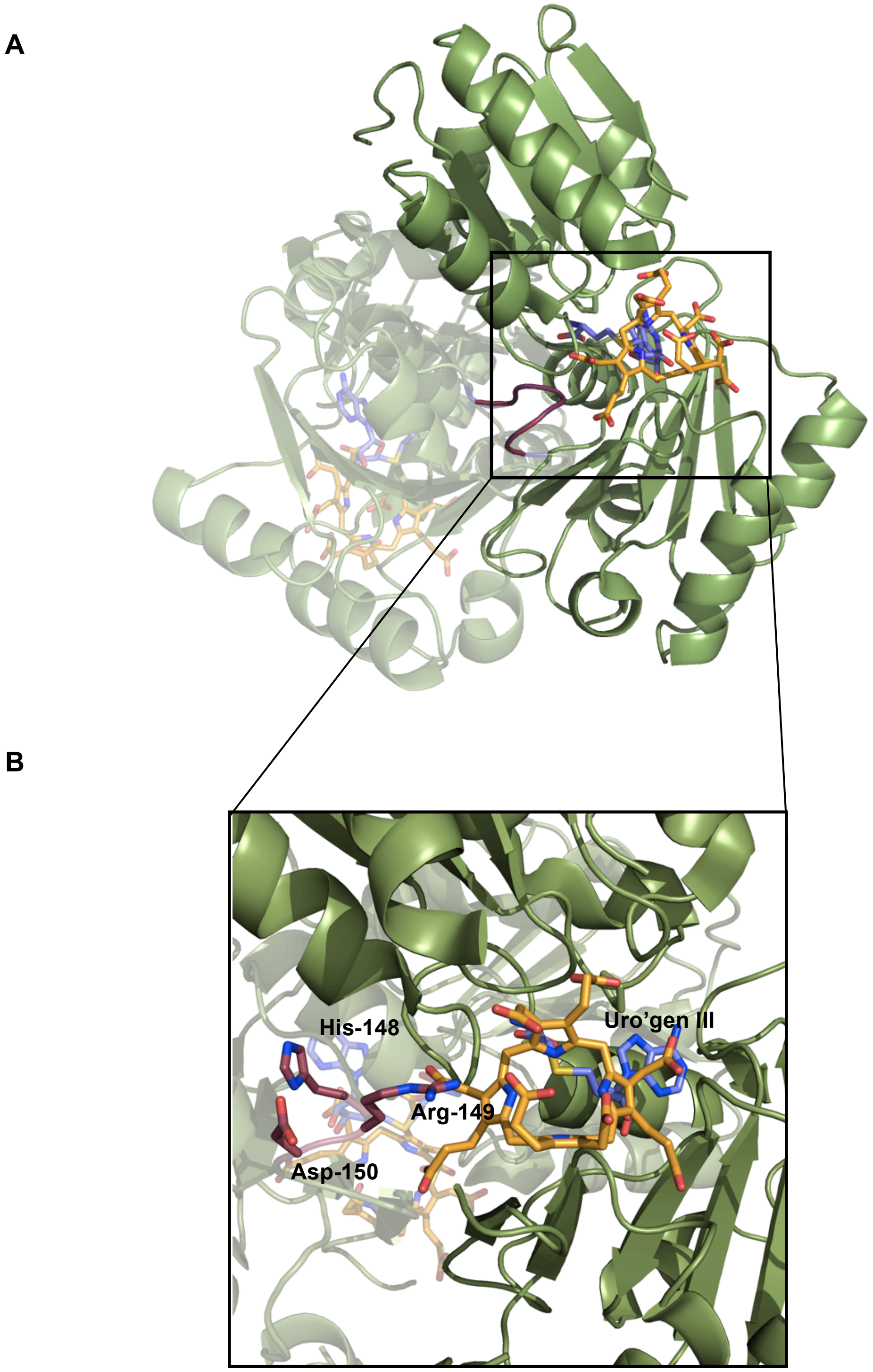


Figure 32: NirE dimer showing the flexible loop of domain B. (A) NirE dimer showing the presence of the flexible loop (highlighted in red) of one monomer extending into the active site pocket of the other monomer. (B) Magnified view of the NirE active site showing the arginine residue from the other monomer interacting with the substrate, uro’gen III.

4B.4. Comparison of NirE with its homologues

The highly conserved amino acid residues among SUMTs from different organisms were determined by an amino acid sequence alignment of more than 60 different SUMTs using CLUSTALW (Larkin *et al.,* 2007). The residues which were identical in all SUMTs were identified as highly conserved. The amino acid sequence alignment of NirE with other SUMTs, whose structures have already been published namely, CysG (PDB ID - 1PJQ) from *Salmonella enterica,* CobA (1S4D) from *Pseudomonas denitrificans*, and *Thermus thermophilus* ttSUMT (1V9A) revealed a sequence identity of 49%, 41% and 40%, respectively. Thus, highest sequence identity was observed between NirE and CysG. A structure based sequence alignment of these four SUMTs was performed using ESPRIPT (Gouet *et al.,* 1999) (Figure 33). The highly conserved amino acids could be mapped in the sequence alignment (indicated by a black star in Figure 33). The amino acids which were identical in all these four SUMTs are marked by a red box. The figure also highlights the helices and β-strands as found in the structure of NirE. Additionally, the amino acids, which formed the flexible regions that could not be included in the final structure, have been indicated by dashed green lines. Based on this sequence alignment and structural analysis of SAH and uro’gen III in the active site of NirE, several residues were identified for site directed mutagenesis (indicated by a blue triangle in Figure 33).

The structure of NirE was found to be highly similar to the above mentioned SUMTs. The root mean square deviations between common C_α positions are 0.603 Å for NirE and CobA, 0.776 Å for NirE and CysG and 0.836 Å for NirE and ttSUMT. Hence, NirE and Cob A demonstrate the highest degree of structural similarity. Structural comparison of the four SUMTs further revealed the existence of two flexible loops in all the four structures. Figure 34 shows the superposition of NirE with the structures of the three SUMTs in ribbon representation. In the NirE structure, these loops are located between β-3 and α-helix D of domain A and between the β-6 and α-helix G of domain B. In Figure 34, the beginnings and ends of flexible loops are indicated by a star and an arrow head. These two loops have been found to be predominantly constituted of disordered amino acids in all SUMTs. Although the flexible loop of domain B does not exhibit the presence of any highly conserved amino acid residue, the domain A loop harbors a highly conserved lysine (Lys-72 in NirE). The corresponding lysine in CysG (Lys-270) was found to be a functionally essential residue (Stroupe *et al*., 2003). Similar to other

SUMTs, in the structure of NirE-uro'gen III complex, the two flexible loops were still found to be disordered and Lys-72 was not included in the final model. Hence, the exact function of this lysine residue is yet to be ascertained.

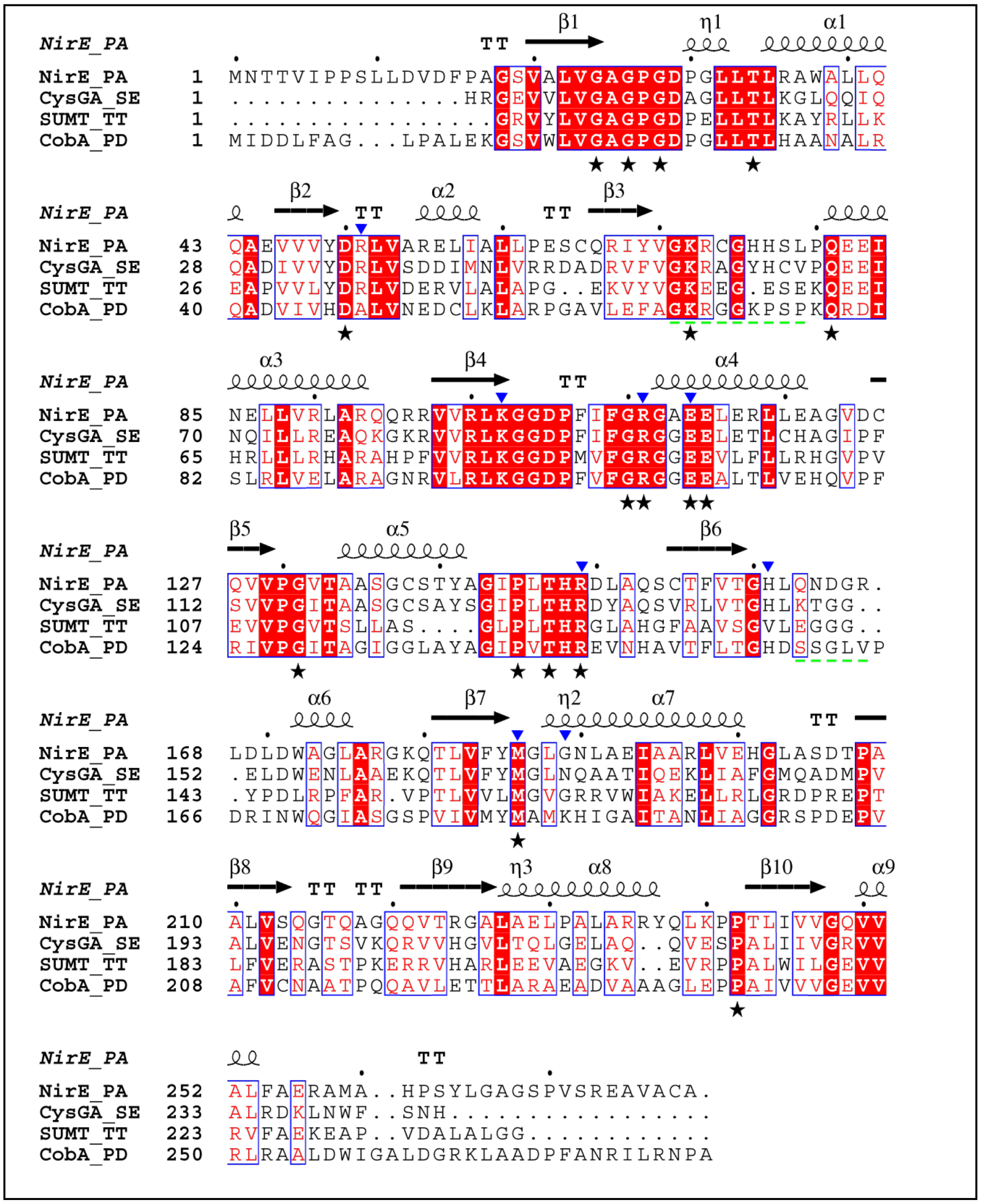


Figure 33. Comparison of NirE with its homologues. Structure based amino acid sequence alignment of NirE from *Pseudomonas aeruginosa* (NirE_PA) with the SUMT domain of CysG from *Salmonella enterica* (CysGA_SE) (Stroupe *et al.,* 2003), a SUMT from *Thermus thermophilus* (SUMT_TT) (Rehse *et al.,* 2005) and CobA from *Pseudomonas denitrificans* (CobA_PD) (Vevodova *et al.,* 2004). Helices and β-strands as found in NirE are indicated. The amino acids that formed the flexible regions and could not be included in the final structure have been indicated by dashed green lines. The identical amino acids among these four SUMTs are highlighted by red boxes, whereas the highly conserved amino acids as identified by an amino acid sequence alignment of 60 SUMTs from different organisms are marked by a black star. The amino acid residues which were subjected to mutational analysis in this study are highlighted by blue triangles. (Figure adopted from Storbeck *et al.,* 2011)

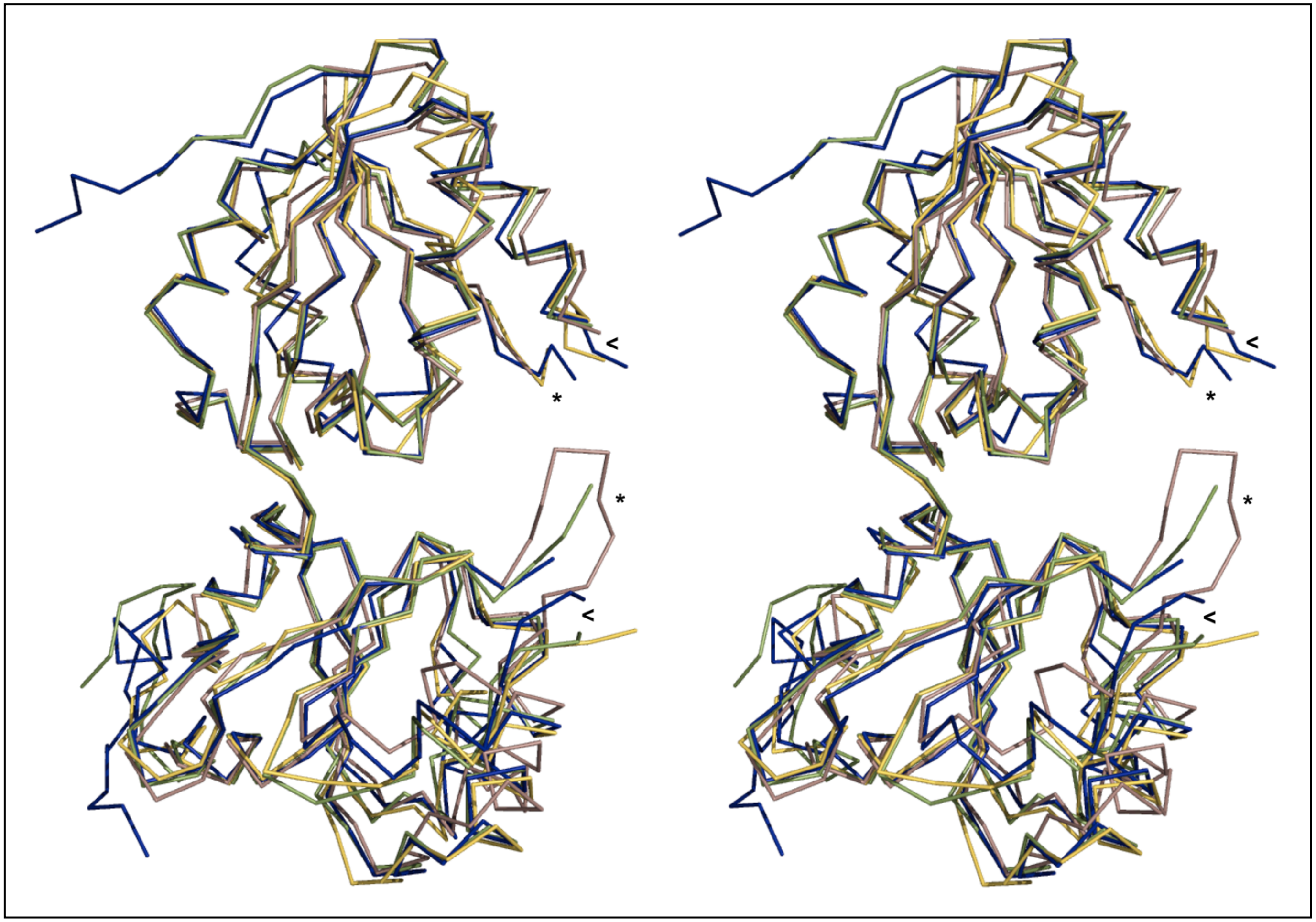


Figure 34: Structural comparison of NirE with its homologues. Superposition of NirE (green) with the structures of CobA (salmon), the SUMT domain of CysG (blue) and ttSUMT (yellow). The structural models are shown in ribbon representation. The beginnings and ends of flexible loops are indicated by a star and a triangle, respectively.

4B.5. Binding of SAH and Uro'gen III at the active site of NirE

SAH was found to bind in a pocket between the two domains of the NirE monomer. In both the structures of NirE, i.e. with and without bound uro’gen III, SAH was buried deep inside this pocket (Figure 35). Consistent with the previously published SUMT structures, the SAH molecule adopts the same bent conformation in both the NirE monomers. The localization and coordination of SAH in both the structures are identical irrespective of the presence or absence of uro'gen III. Figure 35 shows a stereo view of the amino acids involved in the coordination of SAH in the NirE active site. Polar contacts between SAH and the protein mainly involve hydrogen bonds and salt bridges. The amino acids involved in such interactions are Pro-27, Gln-217, Gly-215, Thr-243, Thr-133, Asp-105, Ile-108 and Met-186. The amino acids contributing to the hydrophobic interactions are Leu-52, Tyr-185 and Pro-242. Pro-242 contributes to the hydrophobic environment of the adenine ring of SAH, whereas Tyr-185 stacks along the aliphatic part of the homocysteine moiety. A schematic two-dimensional representation of the interactions between SAH and the active site residues of NirE has been shown in figure 36.

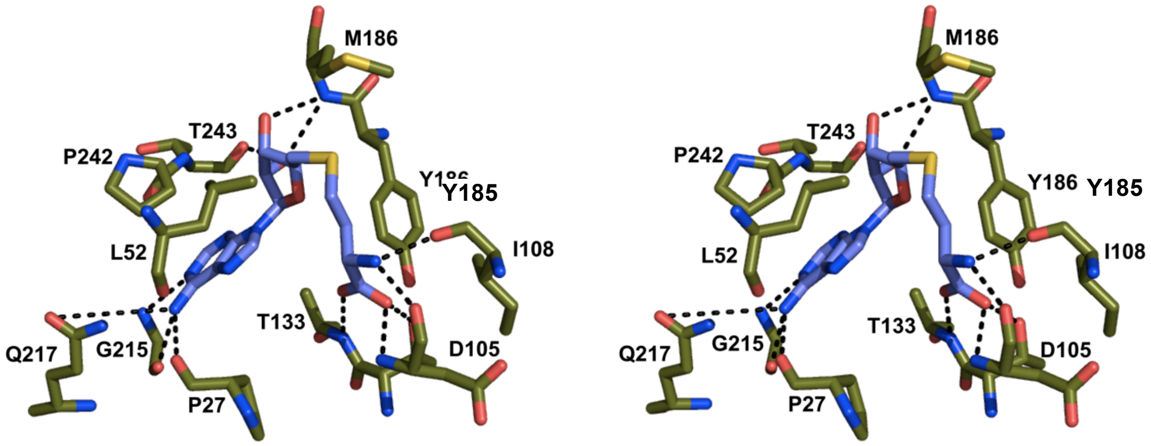


Figure 35: Stereo view of the orientation of SAH in the NirE active site. SAH (blue) along with the amino acids (green) involved in its coordination at the enzyme active site are shown as sticks. The dotted lines represent salt bridges or hydrogen bonds.

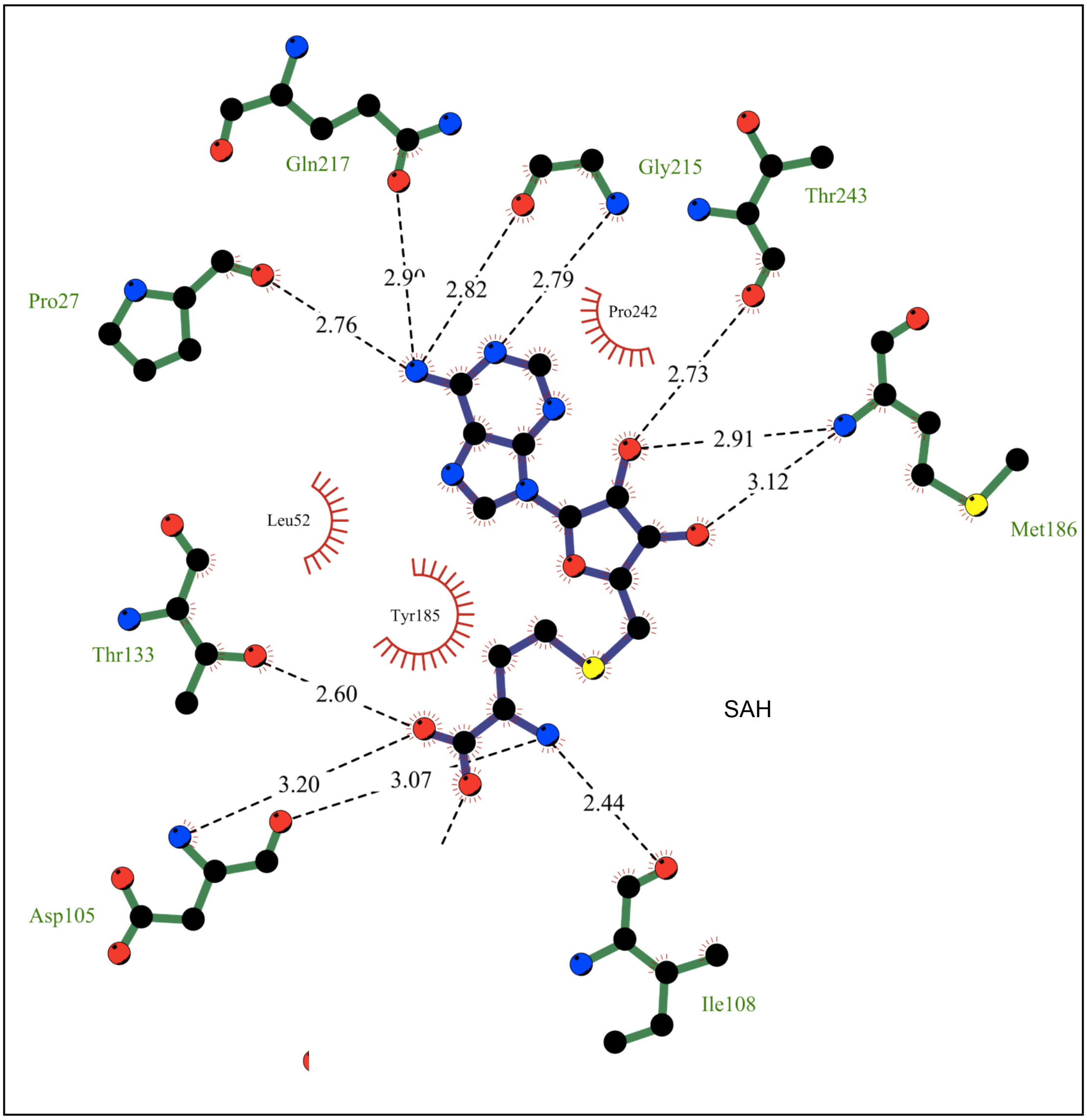


Figure 36: Interaction of SAH with the amino acids in the active site of the enzyme. Schematic two-dimensional representation of the amino acids involved in the coordination of SAH in the enzyme's active site. The dotted line represents a salt bridge or a hydrogen bond with their lengths given in Angstroms. The amino acids involved in hydrophobic interactions are surrounded by a red corona. This figure was created with the program LIGPLOT.

While binding to NirE, uro'gen III adopts a twisted “two-up, two-down” conformation, i.e. the two pyrrole rings facing each other point in the same direction, either upward or downward, out of the tetrapyrrole plane (Figure 36 and 37). Pyrrole rings B and D point upward whereas the rings A and C point down towards the interior of the active site pocket and the bound SAH. Rings B and D are both inclined by about 45°, whereas rings A and C are inclined at an angle of about 70° and 60° out of the tetrapyrrole plane, respectively. The enzyme-product complex structure of uro’gen III synthase also exhibited such a “puckered” conformation (Schubert *et al.,* 2008). Interestingly, Met-186 is located directly below the tetrapyrrole and stacks against pyrrole rings B and D with its sulfur atom being placed beneath the exact centre of the macrocycle (Figure 37A, B). Apart from the interaction between Met-186 and uro’gen III, another hydrophobic contact between the substrate and the protein is provided by the residue His-161. The imidazole ring of His-161 stacks against pyrrole ring C of uro’gen III (Figure 37A).

Polar contacts (hydrogen bonds or salt bridges) between the enzyme and the substrate mainly involve the eight carboxylate groups of the acetate and propionate substituents of uro’gen III with three conserved arginine residues Arg-51, Arg-111 and Arg-149, the NirE protein backbone and several water molecules (Figure 37, 38; water molecules omitted for clarity). The guanidinium group of Arg-111 participates in a polar contact with the acetate carboxylate of ring A. The acetate carboxylate also forms hydrogen bonds with three water molecules, while the propionate carboxylate of this same ring forms a polar contact to the guanidinium group of Arg-51. Additionally, this arginine residue also participates in a polar interaction with the carboxylate group of the acetate substituent of ring B. The ring B propionate side chain is held in its position by hydrogen bonds to the backbone N-H group of Gly-189 and two water molecules. The acetate carboxylate group of ring C forms hydrogen bonds with three water molecules. The propionate carboxylate of the same ring also makes a hydrogen bond with a water molecule. The highly conserved amino acid Arg*-149, which belongs to the polypeptide chain of the second monomer of dimeric NirE (indicated by an asterisk,*) and is located in the flexible loop between α-helix F and β-6 of the C-terminal domain B, also makes an indirect contribution to the coordination of uro’gen III in the active site. A two dimensional representation of these interactions is shown in Figure 38.

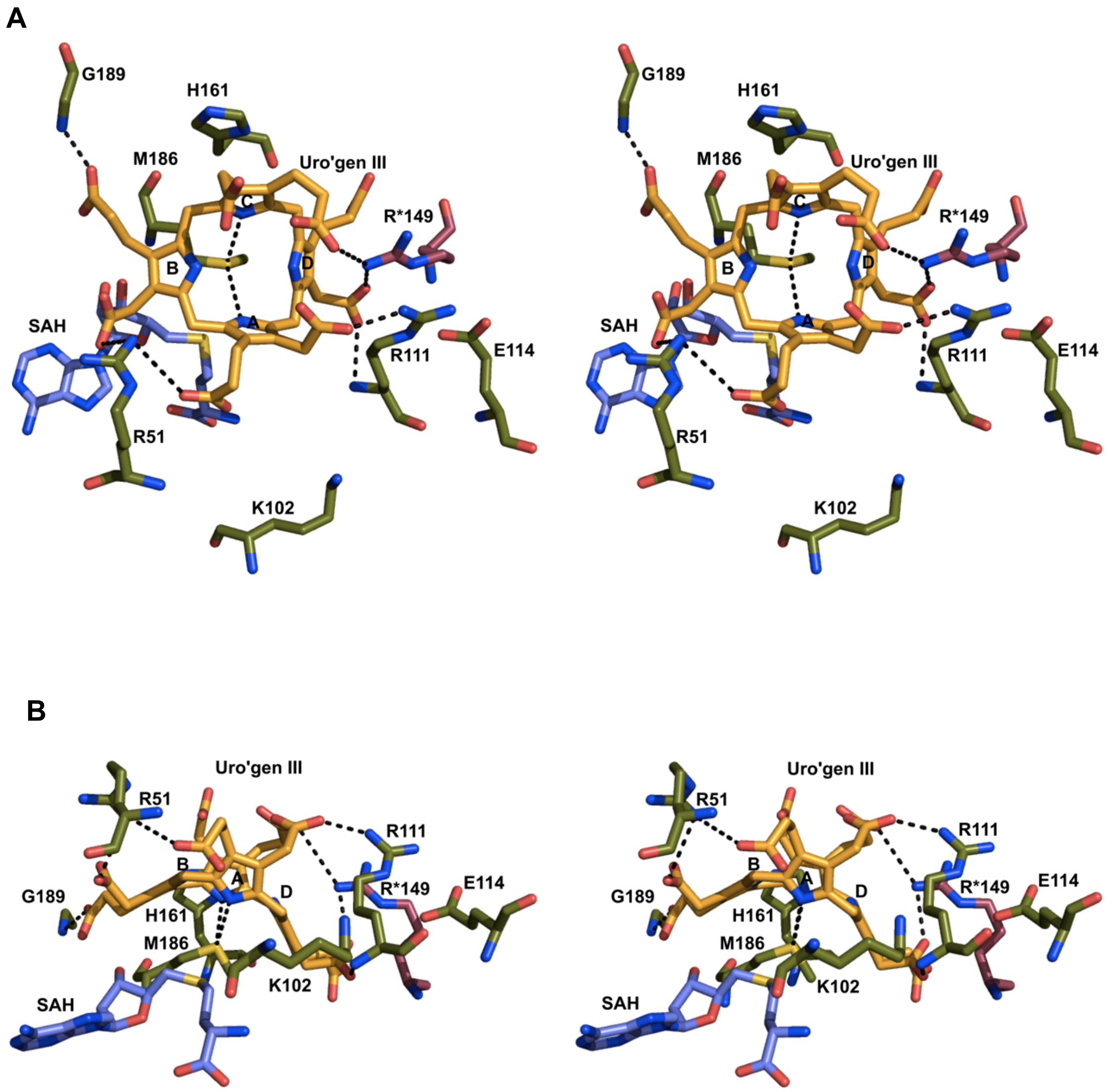


Figure 37: Active site of NirE. (A) Stereo top view and (B) stereo side view of the NirE active site. Important amino acids are shown as sticks. The chain of monomer A is dark green and that of monomer B is purple. The hydrogen bonds and salt bridges are drawn as black broken lines.

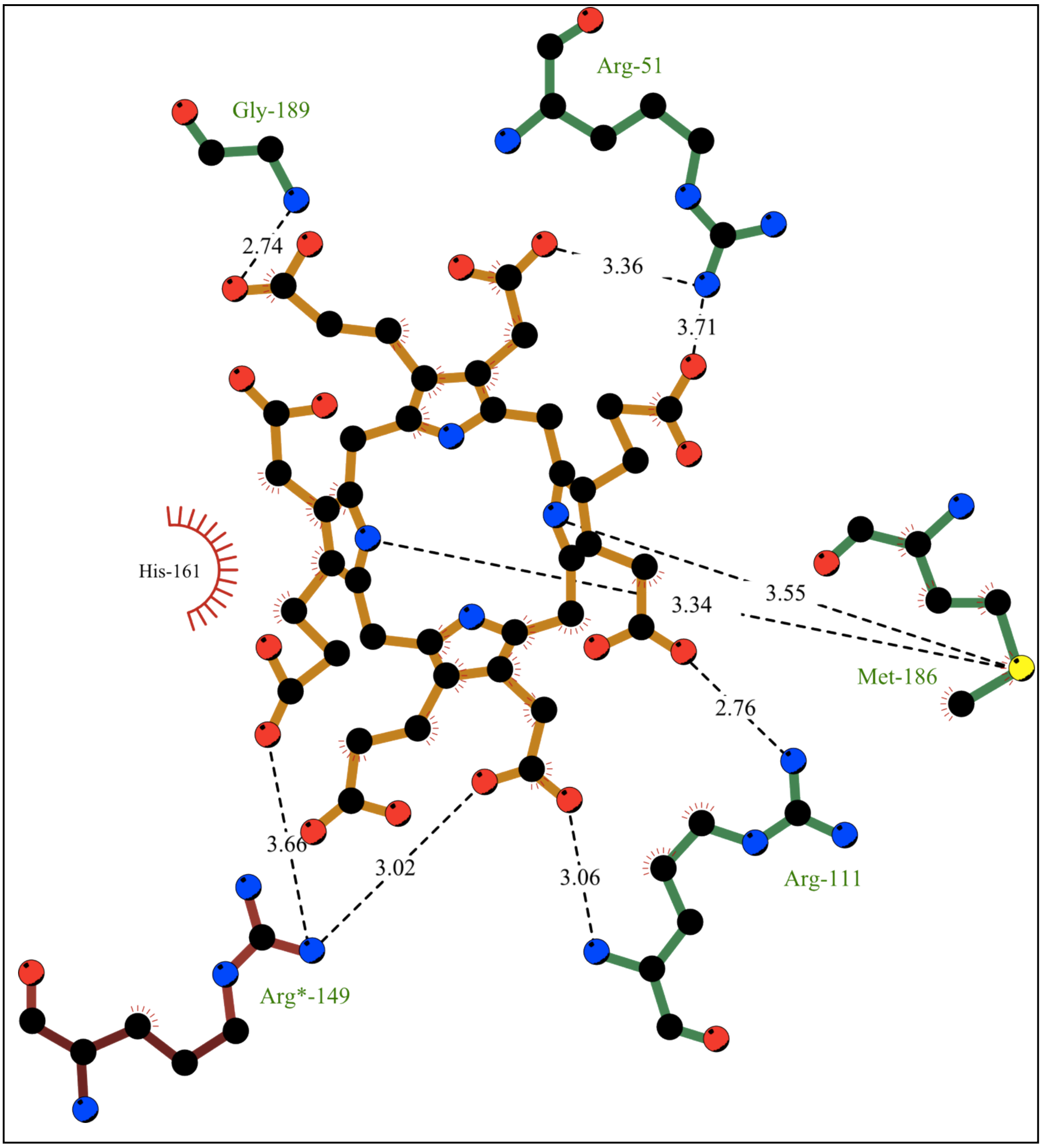


Figure 38: Substrate binding site of NirE. Schematic representation of the interactions between NirE and uro'gen III. Dotted line represents a salt bridge or a hydrogen bond with their lengths given in Angstroms. The Arg*-149 of the other monomer is shown in purple. The red corona around His-161 represents hydrophobic interactions with uro'gen III. This figure was created with the program LIGPLOT.

Of the three highly conserved arginine residues which have been implicated in the coordination of the substrate's carboxylate groups, two residues, Arg-51 and Arg*-149, clearly change their orientation upon substrate binding (Figure 39). In the absence of uro'gen III, the NirE structure shows the guanidinium group of Arg-51 pointing upwards. However, upon substrate binding, it moves down. A significant change in the orientation of the side chain of Arg*-149 is also observed depending on the absence or presence of the substrate. In the absence of bound uro'gen III, this arginine residue makes a salt bridge with the adjacent Asp*-150 in the same monomer. Upon substrate binding, the guanidinium group is flipped to point to the opposite direction towards the active site of the other monomer in dimeric NirE, where it participates in the coordination of the ring C and D carboxylate groups (Figure 39). Another change in the position of amino acids is observed for His-161. This histidine residue is part of the flexible loop of the C-terminal domain of NirE and moves slightly backwards upon substrate binding, in order to accommodate the tetrapyrrole (Figure 39).

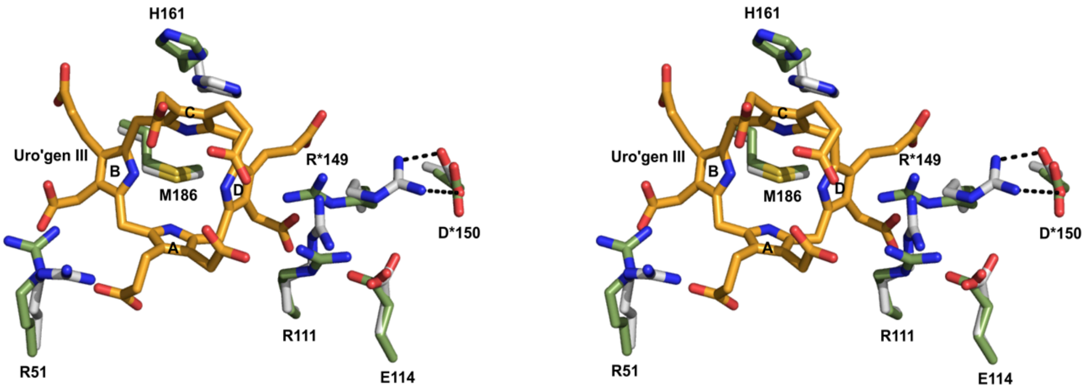


Figure 39: Active site of NirE. Superposition of the NirE active site in the presence (green) and absence (gray) of uro'gen III. Amino acid residues Arg-51, Arg*-149 and His-161 clearly change their orientation upon substrate binding.

4B.6. Generation of NirE variants carrying amino acid exchanges of residues potentially involved in catalysis

The structure of NirE-uro’gen III complex provided an insight into the direct amino acid environment of the substrate uro’gen III in the active site of a SUMT for the first time. This structure revealed the amino acids surrounding uro’gen III that could be involved in substrate binding and/or play a role in catalysis. Based on this structure, amino acid residues, which could have a possible role in catalysis, were identified for site directed mutagenesis. Sonja Storbeck, as a part of her PhD thesis, performed the site-directed mutagenesis. Previously, it has been suggested that a basic amino acid residue is possibly involved in proton abstraction from uro’gen III, that initiates the SAM-dependent methyl transfer reaction (Guillen Schlippe *et al.,* 2005). The possible site of this H^+ abstraction could either be a methylene bridge between the pyrrole rings (most likely the C-20 position for the first methylation), or a pyrrole N-H group. Following the transfer of methyl group from SAM onto uro’gen III, there occurs a rearrangement of bonds within the tetrapyrrole macrocycle (Stroupe *et al.,* 2003, Warren *et al*., 1994). It is possible that the NirE active site’s amino acid residues promote this prototropic rearrangement through acid/base catalysis.

After a detailed structural analysis of the active site of NirE, the following conserved basic residues were selected for site-directed mutagenesis: Arg-51, Lys-102, Arg-111, Arg*-149, and His161. Arg-51 is located near the methylene bridge C-5 between rings A and B, the residues Arg-111 and Arg*-149 are close to the C-20 position between rings A and D, and finally His-161 is located in the direct neighborhood of methylene bridge positions C-10 and C-15 as well as the N-H group of ring C. His-161 was replaced by a phenylalanine. This exchange was done with the aim to knock out the putative function of His-161 as a catalytically active base, while retaining its hydrophobic stacking interaction with pyrrole ring C.

Although arginine is considered to be a weak base, several examples in the literature have previously been reported, demonstrating the involvement of arginine residues in proton abstractions (Guillen Schlippe *et al*., 2005). In these cases, the location of the arginine residues is characterized by their accessibility to solvent molecules and close proximity to the carboxylate groups of the aspartate or glutamate side chains. In the active site pocket of NirE, an acidic environment is created by the highly conserved residues Glu-114 as well as the substrate’s own carboxylate groups, which are close to the conserved arginine residues Arg-51, Arg-111, and Arg*-149.

Hence, all three arginines along with Glu-114, were subjected to site-directed mutagenesis. Arginines were exchanged against lysines and the glutamate was exchanged against glutamine.

Further to the above mutations, Met-186 was also selected for mutational analysis based on its strategic location at the center of the active site and its contribution in binding to both SAH and uro'gen III. In order to investigate the role of this highly conserved methionine, it was replaced by a leucine. This substitution is more conservative than the previously reported methionine to alanine exchange in CobA from *Pseudomonas denitrificans*, which resulted in a complete loss of enzyme activity (Vevodova *et al*., 2004).

Finally, Gly-189 was selected for replacement by an asparagine or a lysine based on the superposition of NirE with the structures of CysG and CobA, in which it corresponds to Asp-385 and Lys-187, respectively. The superposition revealed the potential of Asp-385 and Lys-187 to make polar contacts with the propionate carboxylate of the substrate’s ring B. Hence, the mutagenesis of Gly-189 was designed with the aim to study the positional importance of an asparagine or a lysine in NirE.

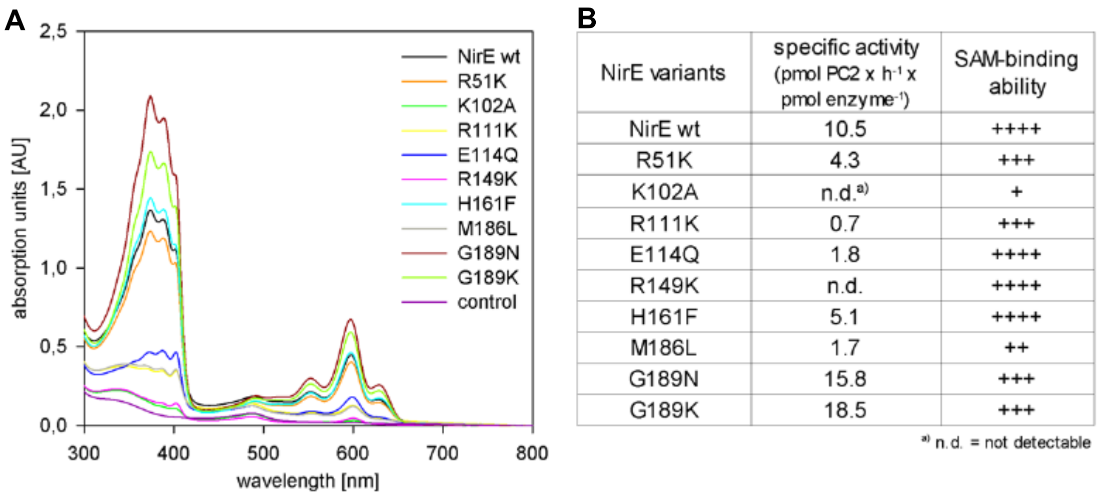
All NirE variants were expressed in *E. coli*, purified, and analyzed for their ability to bind SAM and for catalytic SUMT activity by Sonja Storbeck as a part of her Ph.D. thesis. A summary of the results obtained by her is shown in Figure 40A and B.

4B.6.1. Exchange of amino acid residues Arg-51, Arg-111, Glu-114, Arg*-149, His-161 and Gly-189 does not disturb SAM binding

The binding affinity of SAM to NirE variants G114Q, R149K, and H161F remained unchanged compared to the wildtype protein. A slightly reduced (about 75 % of wt) SAM binding was observed in the case of the NirE variants R51K, R111K, G189N, and G189K. The protein variant M186L exhibited a significant decrease in SAM binding (about 40 % of wt) and the K102A variant demonstrated a highly reduced ability to bind the cofactor (about 20 % of wt). However, the binding of SAM to the M186L variant was strong enough for performing SUMT activity assays and investigating the role of Met-186 in uro’gen III binding and NirE catalysis. The reduced affinity for SAM of K102A was unsuspected as it is not present in the immediate neighborhood of the cofactor. Figure 40 B shows the SAM binding ability of wildtype NirE and the NirE variants.

4B.6.2. SUMT activity assay

Two slightly different assay systems were employed to measure the SUMT activity of NirE wildtype and variants. One of them involved the determination of overnight production of sirohydrochlorin (Figure 40A) whereas the latter measured the specific activity (i.e. the initial reaction rates) (Figure 40B). These assays were done by Sonja Storbeck as a part of her Ph.D. thesis.



NirE variants	specific activity (pmol PC2 x h^{-1} x pmol enzyme^{-1})	SAM-binding ability
NirE wt	10.5	++++
R51K	4.3	+++
K102A	n.d.[a)]	+
R111K	0.7	+++
E114Q	1.8	++++
R149K	n.d.	++++
H161F	5.1	++++
M186L	1.7	++
G189N	15.8	+++
G189K	18.5	+++

[a)] n.d. = not detectable

Figure 40: NirE enzymatic activity. (A) UV-Vis absorption spectra of sirohydrochlorin formed in NirE activity assay mixtures containing either wildtype NirE or variants thereof. (B) The specific activity of wildtype NirE and NirE variants was determined as described in Materials and Methods using chemically produced uro'gen III as the substrate at a final concentration of 17 µM. In contrast to the overnight activity assay (panel B) the values for the specific activity represent the initial reaction rates. The given values represent the average of three independent measurements. The SAM binding ability of wildtype NirE and NirE variants was determined as described in Materials and Methods. The strength of SAM binding is represented as identical to wildtype NirE (+ + + +), slightly reduced (+ + +), reduced (+ +) and strongly reduced (+)

Except for G189N and G189K, all analyzed NirE variants exhibited a reduced SUMT activity compared to the wildtype. In the case of the G189K variant, two-fold increase in specific activity was observed. The most pronounced reduction in SAM binding ability was observed for the variants K102A and M186L. Additionally, a complete loss of SUMT activity was seen in the case of Lys-102 when exchanged against an alanine residue. The M186L variant which showed a relatively low affinity of cofactor binding (of about 40% compared to wildtype) retained only a residual amount of catalytic activity. NirE variants R51K and H161F retained their SAM binding ability but demonstrated reduced specific SUMT activities. However, both variants were still clearly able to catalyze the methyl transfer reaction. The overnight incubation assay revealed that the amount of precorrin-2 produced by them almost equaled that of wildtype NirE.

Similar to R51K and H161F, the NirE mutants R111K and R149K are also characterized by an ability to bind SAM, but have reduced enzymatic function. However, the NirE mutants R111K and R149K still had the ability to form precorrin-2, but both of them demonstrated drastically reduced SUMT activity. Interestingly, both arginines, Arg-111 and Arg*-149, are strategically located in close proximity to the uro'gen III C-20 position, which is a possible site of the proton abstraction. Additionally, both arginines are in close contact to several carboxylate groups which are originating either from the substrate or from the protein. This creates an environment conducive for them to act as base for proton abstraction (Guillen Schlippe *et al.,* 2005). The strictly conserved Glu-114 is also in close contact to Arg-111, which facilitates in positioning it near the substrate. To investigate this hypothesis, the conserved Glu-114 was exchanged with a glutamine. This resulted in a NirE E114Q variant with a SAM binding ability similar to the wildtype NirE, but with significantly reduced SUMT activity.

5. Discussion

5B. Role of Cnx1E in the synthesis of the molybdenum cofactor

5A.1. Active site of Cnx1E

Structural comparison of Cnx1E sub-domain 3 with Cnx1G revealed a high degree of similarity. Apart from minor structural variations, the overall architecture of the two structures is largely the same. This strongly indicates that that sub-domain 3 harbors the active site of Cnx1E catalyzing the deadenylation of MPT-AMP. Previous structural analysis of the Cnx1G-MPT complex provided a detailed insight into its active site pocket. MPT was found to bind in a large predominantly negatively charged cavity within the compact globular protein (Figure 41A). The amino acid residues involved in MPT binding are located on one side of this cavity whereas the residues involved in catalysis are on the other side (Figure 41A). Three residues Thr-542, Ser-573 and Ser-583 were found to be involved in binding and coordination of MPT whereas the residues essential for catalysis are Ser-476, Asp- 486, Asp-515, Asp-548 (Kuper *et al.*, 2003, Kuper *et al.*, 2004).

Similar to Cnx1G, the sub-domain 3 of Cnx1E was also found to possess a negatively charged cavity on one side of the molecule (Figure 41B). The cavity is however much larger in the case of Cnx1E. Three out of four catalytically important aspartate residues (Asp-486, Asp-515 and Asp-548) of Cnx1G are conserved in sub-domain 3 of Cnx1E. The corresponding residues in Cnx1E are Asp-217, Asp-200 and Asp-274 (Figure 41B), of which Asp-274 has already been established as a functionally relevant residue (Corinna Probst, *Master's Thesis*, 2012). Previous studies have shown that molybdate and MPT-AMP bind cooperatively to Cnx1E. Hence, the high positive charge of the surface immediately next to the negatively charged cavity must be the reason to facilitate the binding of the negatively charged molybdate. It is already known that the binding of molybdate at the sub-domain 3 of Cnx1E is essential for the insertion of molybdenum into MPT. The hydrolysis of MPT-AMP and insertion of molybdate into MPT take place subsequently. Hence, on the basis of this comparative analysis of the structure of the active site pocket of Cnx1G to the cavity observed on the surface of domain 3 of Cnx1E, it can be concluded that MPT upon being adenylated by Cnx1G is received by the sub-domain 3 where it gets further processed to yield mature Moco.

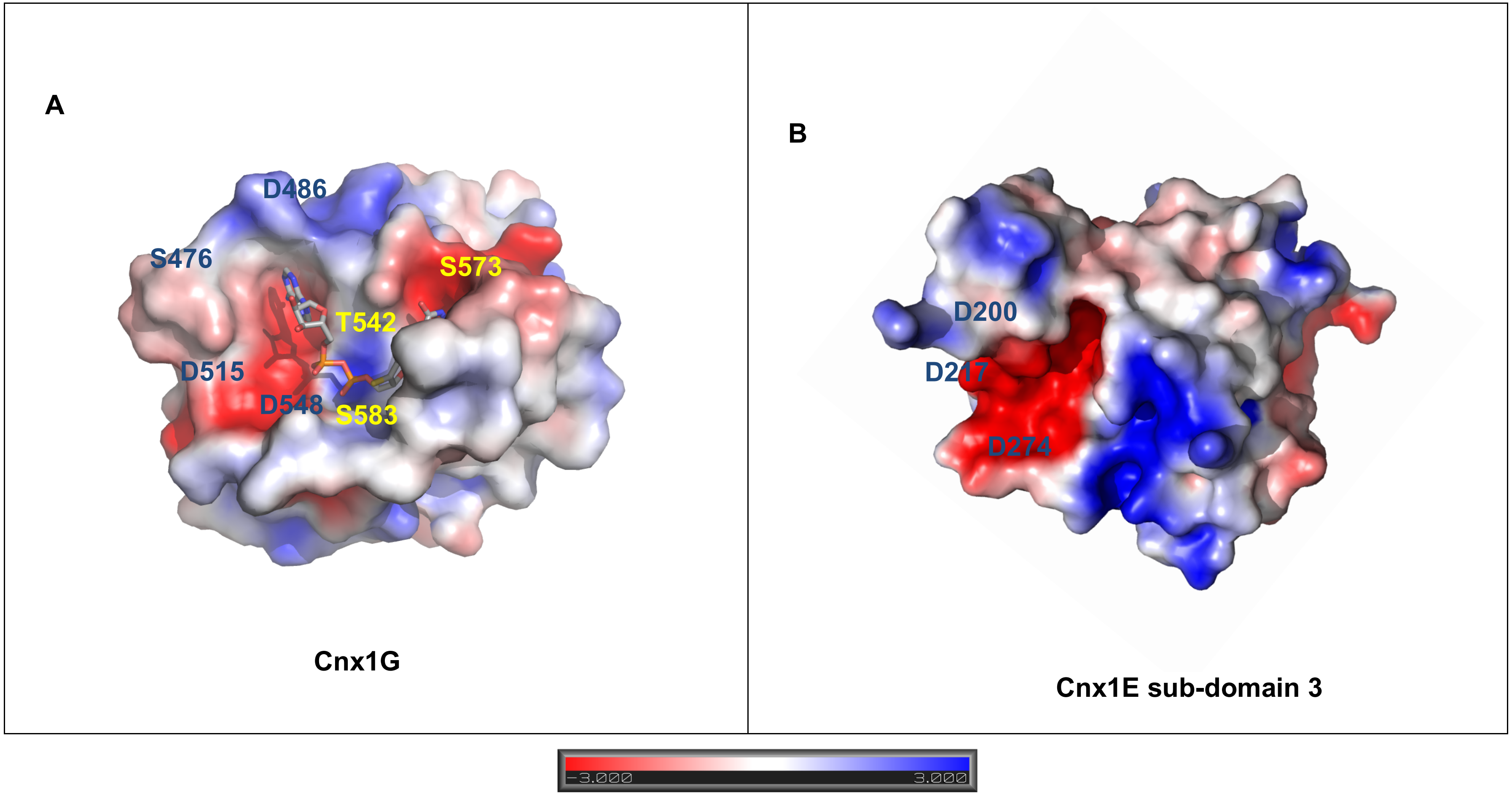


Figure 41: Surface representation of the active site of Cnx1G and Cnx1E putative active site. (A) Negatively charged active site pocket of Cnx1G. Amino acids involved in substrate coordination were mapped on one side of this cavity (yellow) whereas the residues involved in catalysis were mapped on the other side (blue). (B) Negatively charged putative active site pocket observed on the surface of the domain 3 of Cnx1E with the conserved corresponding amino acids of Cnx1G that are implicated in catalysis.

5A.2. Cnx1E: A nudix hydrolase

Cnx1E can be classified as a ‘nudix hydrolase’, based on its functional role in catalyzing the hydrolysis of MPT-AMP to yield Moco. Nudix hydrolases is a subfamily within the enzyme superfamily of pyrophosphate hydrolases. The name of this superfamily is derived from their substrates as they catalyze the hydrolysis of pyrophosphate bond of **nu**cleoside **di**phosphate linked to another moiety (**X**), hence the name **nudix**. Hydrolysis reaction of these enzymes is dependent on divalent cations that in most of the cases are Mg^{2+}. A prominent example of a nudix hydrolase is MutT from *E. coli*. that catalyzes the hydrolysis of the base analogue 8-oxo-dGTP. MPT-AMP represents a nudix substrate with the adenosine linked to the MPT moiety by two phosphate bonds and thus, closely resembles the substrate of MutT, which is 8-oxo-dGTP (Figure 42A, B). Additionally, the catalytic function, Mg^{2+} dependent MPT-AMP hydrolysis of Cnx1E is also similar to MutT.

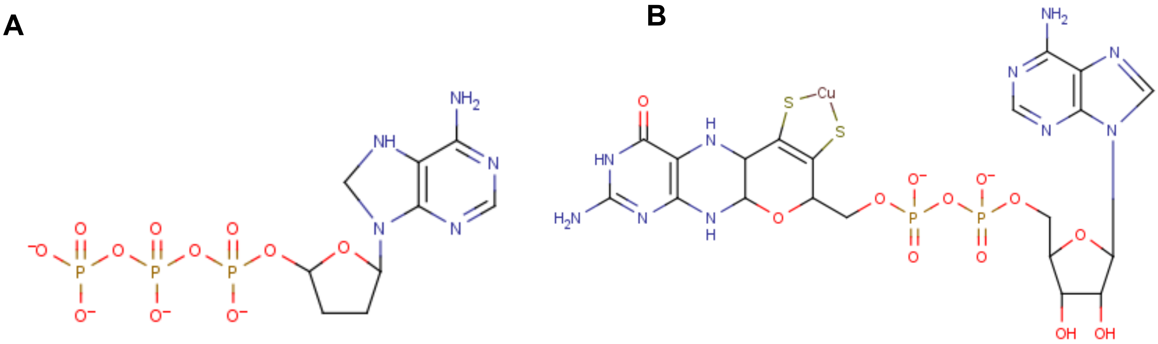


Figure 42: Chemical structures of the substrates of nudix hydrolase. (A) 8-oxo-dGTP, a substrate of MutT from *E.coli*, (B) MPT-AMP

Comparative structural analysis of nudix hydrolases with Cnx1E further asserted its classification in this enzyme family. Nudix hydrolases are characterized by a conserved amino acid sequence known as the nudix motif or the nudix box, $GX_5EX_7REUEXEEXGU$, where X is any amino acid and U is a bulky aliphatic hydrophobic residue. The nudix box folds into a loop helix loop motif, the core of which is mainly formed of glutamic acid residues. These glutamic acid residues play an important role in binding to divalent cations, which in most of the cases are Mg^{2+} ions. Although this loop helix loop motif is not involved in substrate binding, it forms the catalytic site of the enzyme.

The structure of MutT solved in 2010 (PDB ID - 3A6U)(Nakamura *et al.*, 2010), revealed the presence of a loop helix loop motif composed of the nudix box amino acid sequence in the active site. A similar nudix box motif could be mapped to the domain 3 of Cnx1E (Figure 43A, B), which formed the familiar loop helix loop motif (Figure 43D). It contains the α-helix G and the two loops joining this helix to the β-strands β-12 and β-13. This helix is slightly shorter than the helix of the MutT nudix motif (Figure 43C). Although the amino acid residues exhibit less significant sequence identity to nudix box consensus sequence, they show a significant identity amongst each other. In addition to this, the functionally relevant amino acid Asp-274 is part of this motif. This further reinforces the hypothesis of sub-domain 3 harboring the active site of the enzyme.

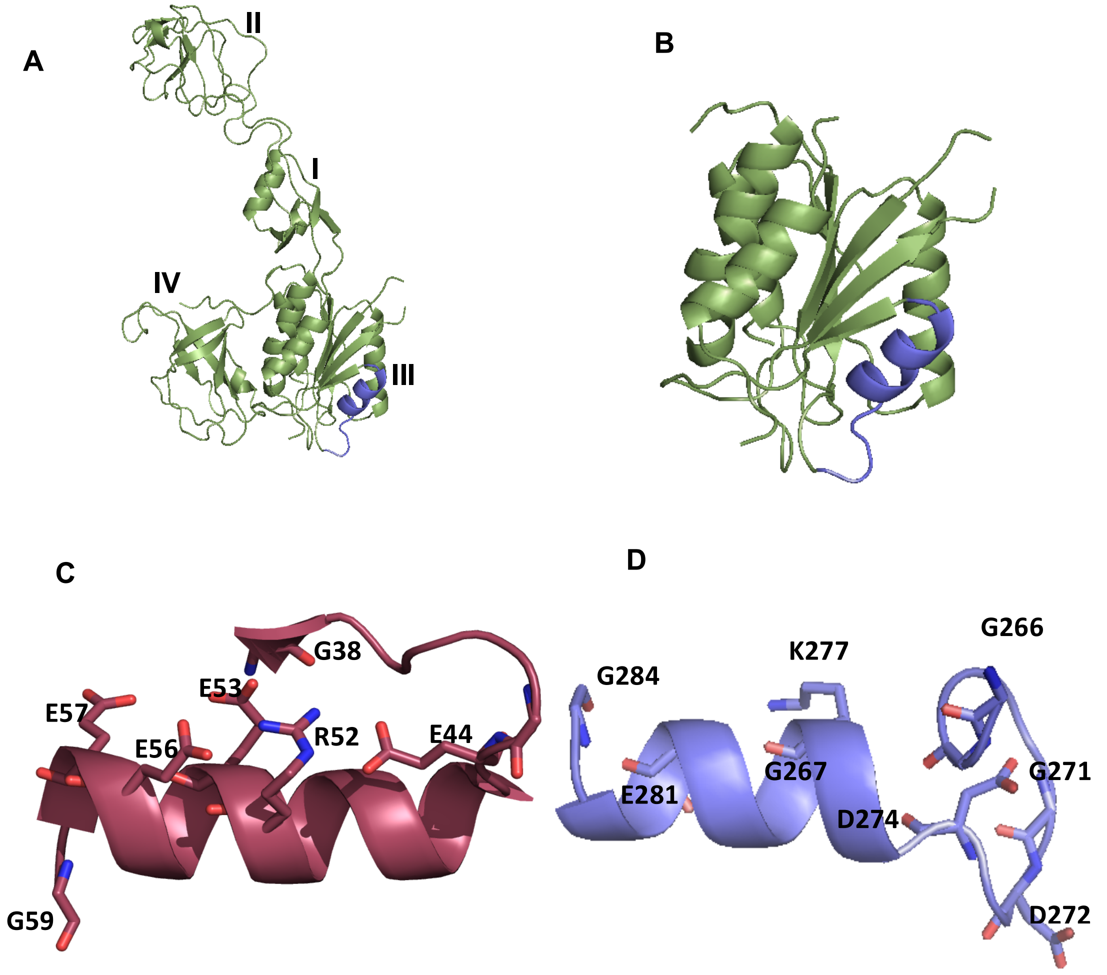


Figure 43: Structure of the putative Nudix box motif of Cnx1E and its location in the tertiary structure. (A) The Cnx1E monomer with the nudix motif colored in blue. The individual sub-domains are denoted by roman numbers alongside. (B) Detailed view of domain3 of Cnx1E showing the location of the nudix motif, (C) the loop-helix-loop motif of the nudix hydrolase which contains the nudix box, (D) Putative nudix motif of Cnx1E.

5A.3. Functional characterization of residues implicated in catalysis and substrate binding

The presence of bound MPT-AMP in all the four cases confirmed that the mutants were not completely active. On comparing the MPT-AMP content of the four mutants with each other, it is seen that the introduction of an additional G108D mutation results in a partially active enzyme as indicated by the reduced MPT-AMP content. This is contradictory to previous studies where G108D mutation was seen to increase the MPT-AMP content of the enzyme. Additional experiments based on the analysis of molybdate saturation in Cnx1E further confirmed the above results. On comparing the molybdate saturation with determined amounts of MPT-AMP bound to the same variants, it is seen that the molybdate or MPT-AMP matches saturation ratios of the four variants. Hence, the direct relation of cooperative binding of molybdate and MPT-AMP to Cnx1E was clearly established. On this basis, we can determine MPT-AMP saturation in Cnx1E variants by the colorimetric determination of molybdate saturation.

In addition to the above experiments, the MPT-AMP hydrolysis proved that the four Cnx1E variants were indeed MPT-AMP hydrolysis mutants. Hence, based on the results of biochemical characterization and functional assays, it is concluded that Asp-274 which is located in the sub-domain 3 of Cnx1E is crucial for biocatalytic property of the enzyme.

5A.4. Implication of Asp-274 on the enzymatic mechanism

the structural analysis and biochemical results, the potential mechanism MPT-AMP hydrolysis involving the role of the amino acid Asp-274 in Cnx1E can be proposed. *E. coli* Nudix hydrolase MutT was used as a model for the functional assignment of the amino acid to the reaction mechanism of the Asp-274 in Cnx1E. MutT catalyzes a nucleophilic substitution (SN). The mechanism is shown in Figure 4.2 A.

As seen in all Nudix hydrolases, the catalytic activity of MutT is also based on a Nudix box. In MutT, the amino acids of the nudix box have been entrusted with two functions. Firstly, they are responsible for the coordination of the Mg^{2+}, which is involved in the catalysis of the hydrolysis reaction. The second task of the Nudix box amino acids is the formation of the actual nucleophile. The amino acid Glu-53 acts as a general base and abstracts a proton from a water molecule. This results in the weakening of the OH bond of the water, thereby increasing the negative charge of the water molecule which favors the nucleophilic attack by this water molecule on the phosphate (Mildvan et al., 2004).

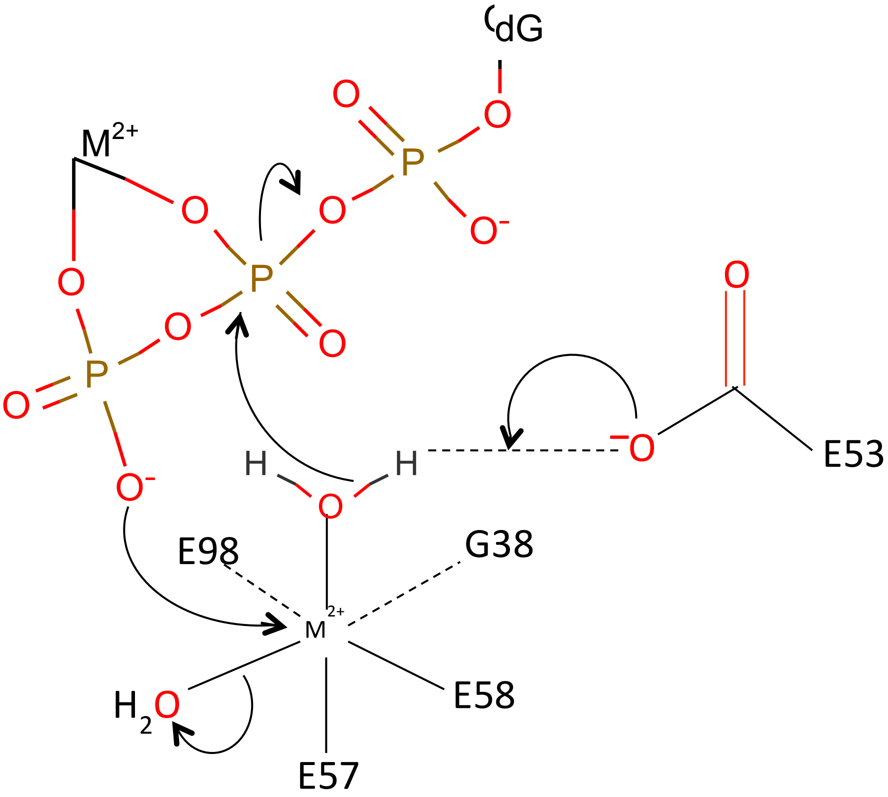


Figure 44: Reaction mechanism of MutT. Reaction mechanism of MutT showing the hydrolysis of phosphate bond catalyzed by Glu-53.

Based on the functional characterization and the location of the amino acid Asp-274 in the putative catalytic site, a speculation could be made. It is possible that Cnx1E Asp-274 is the corresponding residue to Glu-53 of MutT. Hence, Cnx1E Asp-274 might be involved in a nucleophilic attack on a water molecule which would then be available for attacking the phosphate bond between MPT-AMP. However, in the absence of the structure of an enzyme-substrate complex it is impossible to arrive at a definite conclusion. Hence, this needs further investigation.

5B. Role of NirE in the synthesis of the tetrapyrrole Heme d_1

5B.1. Active site of NirE

The high degree of structural similarity of NirE to the previously published SUMTs, suggested that NirE shares similar active site architecture with other members of this enzyme family. This belief was further supported by the location of the catalytic pocket between the two domains of each monomer, a characteristic structural feature of SUMTs (Stroupe *et al.,* 2003, Vevodova *et al.,* 2004, Rehse *et al.,* 2005). A remarkable similarity is observed in the coordination of the SAH molecule in the active site of NirE and CysG, which apparently also exhibits the highest degree of amino acid sequence identity to NirE compared to the two other crystallized SUMTs. The acetate and propionate groups were unambiguously distinguishable from the size and shape of the electron density for these substituents and hence it was possible to assign the respective pyrrole rings as rings B or D. This presented two possible orientation of the uro'gen III molecule which could be inter-converted by a 180° over-head flipping action of the molecule while keeping the positions of rings A and D unchanged. However, on the basis of the known stereochemistry of the methylation reaction and the location of the SAH molecule beneath one face of uro'gen III, only one of the two orientations was possible. In the structure of NirE in complex with uro'gen III, the orientation of the substrate shows rings A-D arranged in a clockwise manner when viewed from top. The central part of the substrate is stabilized by the residue Met-186 that provides a hydrophobic environment and forms hydrogen bonds (~3.5 Å each) with the N-H groups of pyrrole rings A and C (Figure 38). The angles and distances between the Met-186 sulfur and the A and C rings' N-H groups are in accordance with the presence of these unusual sulfur-containing H-bonds (Zhou *et al*., 2009). Therefore, it could be rightly said that Met-186 is important for the binding of both SAM and uro'gen III. The role of this methionine in substrate binding has also been experimentally observed in a *Pseudomonas denitrificans* CobA variant in which the corresponding methionine residue (Met-184) on being replaced by an alanine was unable to bind SAM and showed almost no SUMT activity (Vevodova *et al*., 2004).

Overall, the coordination between the acetate and propionate substituents of uro'gen III and the surrounding amino acid residues is quite loose and the substrate is not very tightly fixed. A positively charged environment is provided by the arginine residues Arg-111, Arg-51, and Arg*-149 and possibly also by the highly conserved Lys-72 lying on the disordered loop of domain A. This environment accommodates the carboxylate groups of the substrate. This loose coordination of the substrate's carboxylate groups is reflected in a relatively high temperature factor (58.8 Å^2 compared to the average temperature factor of 34.4 Å^2) and a weaker electron density for the substrate compared to that of SAH (Figure 20A).

The fact that the enzyme has to bind the mono-methylated precorrin-1 in the same active site pocket as that of uro'gen III in order to perform the second methylation at C-7 justifies this relatively loose enzyme-substrate coordination. It further explains the ability of SUMTs to accept uroporphyrinogen isomers other than uro'gen III, such as uro'gens I, II, or IV, as substrates (Warren *et al*., 1990, Scott *et al*., 1989).

5B.1.1 Lys-102 and Met-186 are the two key residues that are primarily involved in SAM binding

As expected from structural analysis, the results of the SAM binding assay, showed that the residues Arg-51, Arg-111, Glu-114, Arg*-149, His-161, and Gly-189 do not play a crucial role in binding SAM. Hence, any decrease in SUMT activity as a result of their substitution should be implicated to their involvement in the coordination of uro'gen III or NirE catalysis. The M186L variant was expected to demonstrate a reduced SAM binding ability as the structure of NirE as well as CobA and CysG structures had revealed the hydrophobic interaction between the SAH and this methionine residue. One surprising finding from this assay was the highly reduced SAM binding ability of K102A as it is not present in the immediate neighborhood of SAM. However, this lysine residue participates in an extended network of hydrogen bonds involving the residues Glu-115, Gly-110, Arg-100, Gln-81, and Asp-50. Except Arg-100, all these amino acid residues are highly conserved in SUMTs from different organisms. Hence, it could be concluded that Lys-102 along with the amino acid residues mentioned above contribute to the overall structural integrity of the active site.

5B.1.2. Amino acids Arg-51, His-161, and Met-186 are involved in uro'gen III coordination.

The primary amino acids involved in uro'gen III coordination are Arg-51, His-161, and Met-186. This is clearly seen in the structure of NirE-uro'gen III complex and was experimentally proved by the enzyme assay. It was expected that the NirE variants in which amino acids had been replaced that play a role in substrate binding but have no direct contribution in catalysis, would demonstrate a decrease in specific activity but an equal overnight activity compared to the wildtype NirE. Both R51K and H161F mutants were observed to retain their ability to methylate uro'gen III and form precorrin-2. Hence, the reason for their reduced specific activity was attributed to their reduced substrate affinity that resulted in impaired substrate coordination at the active site. Initially, these two amino acids were also considered suitable to act as base for proton abstraction from uro'gen III. However, the retention of the catalytic activity in these two mutants ruled out that hypothesis.

Met-186 and Lys-102 are the two residues that are involved in both SAM and uro'gen III binding. The significant decrease in SAM binding ability coupled with reduced SUMT activity reinforced their role in substrate coordination. K102A variant was marked by a complete loss of catalytic activity which could possibly be attributed to a disruption in the active site architecture resulted by the replacement of lysine by alanine. However, in the case of M186L, the reduced SAM affinity of 40% compared to wildtype NirE and retention of residual SUMT activity proved its critical role in SAM and uro'gen III coordination, which is also indicated by its strategic location in the active site. This suggests that the sulfur atom of the methionine is possibly involved in correctly positioning the substrate through hydrogen bonds with the N-H groups of pyrrole rings A and C.

The sharp increase in the activity of the NirE variants G189N and G189K could be attributed to a more precise and specific binding of uro’gen III and precorrin-1 through polar contacts between the asparagine or lysine side chains with the carboxylate groups of the substrates. There is a marked similarity in the active sites of the NirE variants G189N and G189K and those of CysG and CobA, respectively. Interestingly, these two enzymes also exhibited a higher specific activity than NirE in previous studies (Stroupe *et al.*, 2003, Vevodova *et al.*, 2004).

5B.1.3. Amino acids Arg-111, Glu-114 and Arg*-149 are involved in NirE catalysis.

Mutational analysis clearly established the direct involvement of Arg-111, Arg*-149 and Glu-114 in NirE catalysis. The biochemical results strongly suggest that both the arginines are suitable to act as a base in order to abstract a proton from Uro'gen III. The positive charge of the Arg-111 is further enhanced by Glu-114 by forming a polar contact with it. Hence, it can be rightly concluded that Arg-111, Arg*-149 and Glu-114, in addition to the coordination of the substrate's carboxylate side chains, they have a direct part in catalysis. However, it is difficult to state which of the two arginines, Arg-111 or Arg*-149 act as the base to initiate proton abstraction

Both the arginines are equally suited for the role of the base owing to their polar contacts with carboxylate groups. The substrate's ring A acetate side chain and the carboxylates of Glu-114 form polar contacts with Arg-111 whereas in the case of Arg*-149, such polar contacts are furnished by the substrate's ring D acetate side chain and ring C propionate substituent. It has been previously proposed that such a carboxylate-rich environment is a common motif for some proteins that use arginine residues as the catalytically essential base (Guillen Schlippe *et al*., 2005). The role of carboxylate in such reactions has been stated to facilitate the rapid exchange of protons with the solvent. Additionally, solvent accessibility of their active sites is another requirement of this group of enzymes. This feature is clearly seen in the case of NirE from the structure of NirE-uro'gen III complex.

On considering the distances of the two arginine residues from the C-20 position, there seems a slight advantage of Arg*-149 over Arg-111. In the case of Arg*-149, the closest guanidinium NH_2 group to C-20 is at a distance of 4.66 Å whereas for Arg-111 this distance is 4.75 Å. The terminal amide group of this residue is located at a distance of 6.15 Å to C-20. Since SAH is bound to the NirE-substrate structure, which is devoid of the methyl group, it allowed for a deeper positioning of the uro'gen III than would have SAM. Hence, in the presence of the true co-substrate, SAM in the active site, the substrate's C-20 position would move even closer to the both arginine residues. Such competing facts could also mean that both Arg-111 and Arg*-149 are required for catalysis. Another striking observation that might explain this hypothesis is their direct neighborhood to each other. In fact, the two guanidinium groups are arranged almost parallel to each other with a distance of about 3.5 Å. This close proximity creates the opportunity for the positive charge of one of the guanidinium groups to favor the deprotonated state of the other which would then be available for proton abstraction from C-20. In the absence of the substrate, Arg*-149 forms a salt bridge with Asp*-150 in the same monomer. Upon substrate binding, the side chain of Arg*-149 moves away from Asp*-150 towards the active site of the other monomer of the NirE dimer where it participates in polar contacts with two carboxylate side chains of the substrate. This new position of Arg*-149 brings it closer to the Arg-111 residue that favors the deprotonation of the former. Subsequently, uncharged Arg*-

149 can then act as the base and abstract a proton from C-20. However, the reverse of this theory is also possible.

5B.2. Implications for the catalytic mechanism of NirE and other SUMTs.

The crystal structure of NirE in complex with its substrate uro’gen III presented an opportunity to identify a set of amino acid residues that possibly bring about the catalytic transformation of uro’gen III to precorrin-2. Further investigation based on biochemical analysis of the functional role of these residues on NirE enzyme activity narrowed down this list to three amino acids: the two arginines, Arg-111 and Arg*-149 and the Glu-114, all of which have a direct involvement in catalysis. All these three residues are highly conserved among different SUMTs from various organisms. On the basis of localization of Arg-111, Arg*-149 and Glu-114 with respect to the substrate uro’gen III observed in the crystal structure, a potential mechanism for NirE could be postulated, which can very rightly be applied to all members of the SUMT family (Figure 45). Previously, it has already been proposed that the methyl transfer reaction in SUMTs is initiated by a proton abstraction from the substrate uro'gen III (Guillen Schlippe *et al.,* 2005). On the basis of structural information obtained from the enzyme- substrate complex of NirE-uro’gen III and functional characterization of NirE variants, it is concluded that either Arg-111 or Arg*-149 acts as the catalytically active base and is responsible for proton abstraction from C-20 of uro’gen III thereby facilitating the nucleophilic attack of C-2 at the methyl group of SAM.

Structural analysis of the active site of the enzyme-substrate complex revealed that that puckered conformation of uro’gen III in the enzyme active site, results in closer positioning of the N-H group of ring A to the sulfur atom of SAH while leaving the C-2 atom at a relatively large distance to the SAH sulfur (6.15 Å). To investigate the compatibility of this puckered conformation of uro'gen III in the presence of SAM in the active site, SAH was replaced with SAM by *in silico* modeling under the assumption that it would adopt an identical conformation. The basis of this assumption was the crystal structures of SAM complexes (e.g. PDB ID 3GX3 and 3GX6, (Montange *et al.*, 2010)) that have been previously published. From the model, it was observed that the puckered conformation of uro’gen III in the presence of SAM would result in a clash between the N-H group of ring A with the methyl group of the co-substrate (Figure 46A). Hence, it is concluded that, during actual catalysis when SAM is present in the active site, uro'gen III binds in a conformation in which the ring A is flipped down in order to avoid this clash (Figure 46B). This conformation of uro’gen III with a planar arrangement of ring A potentially resembles that of the reaction transition state and might facilitate the formation of the new double bond between C-20 and C-1 upon proton abstraction from C-20. The flipping of the ring A also aids in positioning the C-2 atom closer to the methyl group of SAM that facilitates the nucleophilic attack. Subsequent to the methyl group transfer, a rearrangement of double bonds occurs within the tetrapyrrole macrocycle. Consistent with this hypothesis, it was observed that a double bond which was originally located between C-3 and C-4 in uro’gen III, is now found either between C-4 and C-5 or between C-4 and N-21 in the mono methylated intermediate precorrin-1 (Brunt *et al*., 1989). This tautomerization of the tetrapyrrole might probably be influenced by the highly conserved Lys-72 of NirE, as previously proposed for the corresponding Lys-270 of *Salmonella enterica* CysG (Stroupe *et al*., 2003). Unfortunately, in the NirE structure, this lysine residue is located on the flexible loop of domain A that could not be resolved. However, it has been reported that the K270I variant of CysG showed negligible SUMT activity (Stroupe *et al.,* 2003) that reinforces the belief that Lys-72 has an essential contribution in catalysis. Additionally, it is possible that the water molecules, which are present around uro’gen III, might also be involved in the tautomerization of precorrin-1. On completion of the first methylation, the mono-methylated intermediate precorrin-1 (Figure 46C) and the reaction by-product SAH are released from the active site. In order to facilitate this release, the salt bridge between Arg*-149 and Asp*-150, which is observed in the absence of the bound uro'gen III is reformed. In order for the second methylation at C-7 to take place, a new SAM molecule needs to bind at the active site pocket followed by the binding of precorrin-1. Precorrin-1 binds to the enzyme in an orientation representing a 90° counter clockwise rotation with respect to the orientation of the initial substrate uro’gen III. As a result of this rotation, the methylene bridge C-5 now occupies the position that was initially occupied by C-20 during the first methylation. Hence, again Arg-111/Arg*-149 (Figure 45) would catalyze the proton abstraction from C-5 to initiate the second methyl group transfer. Finally, another rearrangement of double bonds would complete the synthesis of precorrin-2.

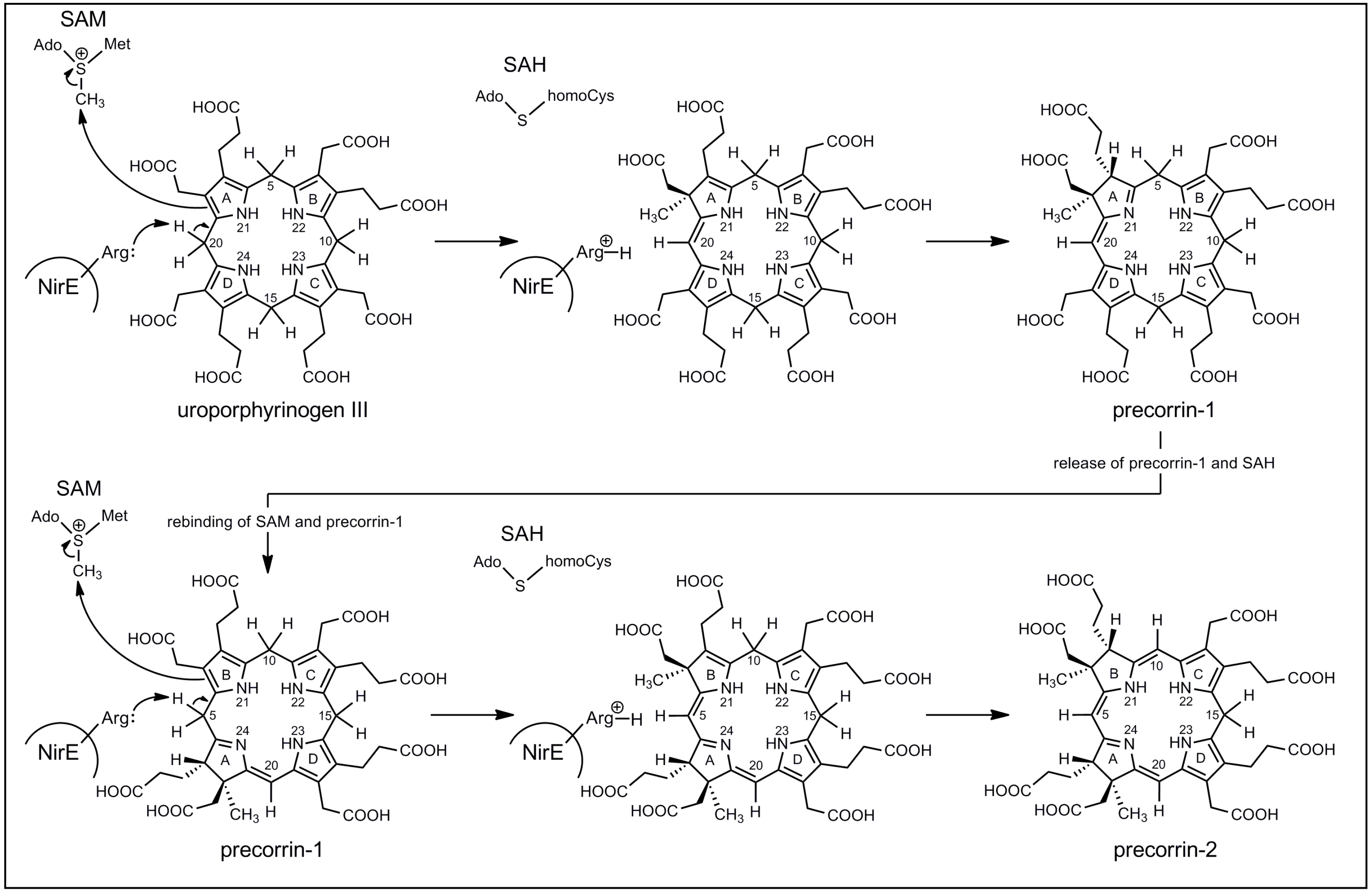


Figure 45. Proposed reaction mechanism of NirE (see text for detailed explanation).

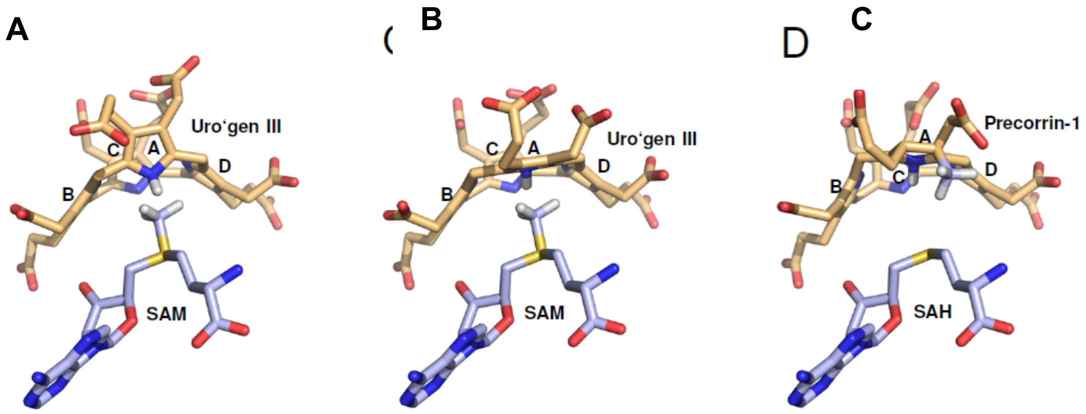


Figure 46. Model of SAM in the NirE active site and proposed conformation of bound uro’gen III in the presence of SAM in the NirE active site. The molecules are shown in stick representation. SAM/SAH is shown in light blue, uro’gen III/precorrin-1 in light orange. (A) Clash between the methyl group of modeled SAM and the N-H group of ring A in case that uro’gen III adopts its “two-up, two-down” conformation. (B) Proposed conformation of uro’gen III in the presence of SAM. (C) Proposed conformation of precorrin-1 after the methyl transfer reaction and prototropic rearrangements.

6. References

6. References

Allen, J.P. & Williams, J.C. (1998). Photosynthetic reaction centers. *FEBS Lett.* **438**, 5-9.

Aucken, C.J., Leeper, F.J. & Battersby, A.R. (1997). Haem d_1: stereoselective synthesis of the reduced form of its parent macrocycle using the original coupling strategy. *J. Chem.Soc.*, *Perkin Trans.* **1**, 2099-2110.

Bali, S., Lawrence, A.D, Lobo, S.A., Saraiva, L.M., Golding, B.T. , Palmer, D.J. , Howard, M.J., Ferguson, S.J. & Warren, M.J. (2011). Molecular hijacking of siroheme for the synthesis of heme and heme d_1. *Proc. Natl. Acad. Sci. U S A.* **108**, 18260-18265.

Bali, S., Warren, M.J. & Ferguson, S.J. (2010). NirF is a periplasmic protein that binds *heme* d_1 as part of its essential role in d_1 heme biogenesis. *FEBS J.* **277**, 4944-4955.

Battersby, A.R., Fookes, C.J.R., Gustafson-Potter, K.E., McDonald, E. & Matcham, G.W.J. (1982). Biosynthesis of porphyrins and related macrocycles. Part 18. Proof by spectroscopy and synthesis that unrearranged hydroxymethylene is the product from deaminase and the substrate for cosynthetase in the biosynthesis of uroporphyrinogen-III. *J. Chem.Soc.*, *Perkin Trans.* **1**, 2427-2444.

Beale, S.I. (1999). Enzymes of chlorophyll biosynthesis*. Photosynthesis Research,* **60**, 43-73.

Beale, S.I. & Castelfranco, P.A. (1973). ^{14}C incorporation from exogenous compounds into 5-aminolevulinic acid by greening cucumber cotyledons. *Biochem. Biophys. Res. Commun.* **52**, 143-149.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D. & Zardecki, C. (2002). The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899-907.

Bertero, M.G., Rothery, R.A., Palak, M., Hou, C., Lim, D., Blasco, F., Weiner, J.H. & Strynadka. N.C. (2003). Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. *Nat. Struct. Biol.* **10**, 681-687.

Bittner, F., Oreb, M. & Mendel, R.R. (2001). ABA3 is a molybdenum cofactor sulfurase required for activation of aldehyde oxidase and xanthine dehydrogenase in *Arabidopsis thaliana*. *J. Biol. Chem.* **276**, 40381-40384.

Brindley, A. A., Zajicek, R., Warren, M. J., Ferguson, S.J. & Rigby, S.E. (2010). NirJ, a radical SAM family member of the *heme* d_1 biogenesis cluster. *FEBS Lett.* **584**, 2461-2466.

Brunt, R.D.,Leeper, F.J., Grgurina, I. & Battersby, A.R. (1989). *J. Chem. Soc. Chem. Commun.* 428-431.

Burgess, B.K. & Lowe, D.J. (1996). Mechanism of molybdenum nitrogenase. *Chem. Rev.* **96**, 2983-3012.

Campbell, W.H. (1999). Nitrate reductase structure, function and regulation: bridging the gap between biochemistry and physiology. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 277-303.

Corinna Probst, Master's Thesis. (2012). Biochemische Charakterisierung von putativen MPT-AMP Hydrolyse Mutanten aus *Arabidopsis thaliana.* TU-Braunschweig.

Crane, B.R., Siegel, L.M. & Getzoff, E.D. (1997). Structures of the siroheme- and [4Fe-4S] containing active center of sulfite reductase in different states of oxidation: heme activation via reduction-gated exogenous ligand exchange. *Biochemistry* **36**, 12101-12119.

Cutruzzola, F., Brown, K., Wilson, E.K., Bellelli, A., Arese, M., Tegoni, M., Cambillau, C. & Brunori, M. (2001). The nitrite reductase from *Pseudomonas aeruginosa*: essential role of two active-site histidines in the catalytic and structural properties. *Proc. Natl. Acad. Sci. U S A.* **98**, 2232-2237.

Cutruzzola, F., Rinaldo, S., Centola, F. & Brunori, M. (2003). No production by *Pseudomonas aeruginosa* cd_1 nitrite reductase. *IUBMB Life* **55**, 617-621.

Dammeyer, T., & Frankenberg-Dinkel, N., (2008). Function and distribution of bilin biosynthesis enzymes in photosynthetic organisms. *Photochem. Photobiol. Sci*. **7,** 1121-1130.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S. & Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucl. Acids Res.* **35**, 375-383.

Dos Santos, P.C., Dean, D.R., Hu, Y., & Ribbe, M.W. (2004). Formation and insertion of the nitrogenase iron-molybdenum cofactor. *Chem. Rev.* **104**, 1159-1173.

Eady, R.R. (1996). Structure-function relationships of alternative nitrogenases. *Chem. Rev.* **96**, 3013-3030.

Eilers, T., Schwarz, G., Brinkmann, H., Witt, C., Richter, T., Nieder, J., Koch., B., Hille, R., Hansch, R. & Mendel, R.R. (2001). Identification and biochemical characterization of *Arabidopsis thaliana* sulfite oxidase. A new player in plant sulfur metabolism. *J. Biol. Chem.* **276**, 46989-46994.

Einsle, O., & Kroneck, P. M. (2004). Structural basis of denitrification. *Biol. Chem.* **385,** 875-83.

Einsle, O., Tezcan, F.A., Andrade, S.L., Schmid, B., Yoshida, M., Howard, J.B., & Rees D.C., (2002). Nitrogenase MoFe-protein at 1.16 Å resolution: a central ligand in the FeMo-cofactor. *Science* **297**, 1696-1700.

Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501.

Enroth, C., Eger, B.T., Okamoto, K., Nishino, T. & Pai, E.F. (2000). Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: structure-based mechanism of conversion. *Proc. Nath.Acad. Sci. USA.* **97**, 10723-10728.

Enroth, C., Schwarz, G., Brinkmann, H., Witt, C., Richter, T., Nieder, J., Koch, B., Hille, R., Hansch, R. & Mendel, R.R. (2001). Identification and biochemical characterization of *Arabidopsis thaliana* sulfite oxidase. A new player in plant sulfur metabolism. *J. Biol. Chem.* **276,** 46989-46994.

Ermler, U. (2005). On the mechanism of methyl-coenzyme M reductase. *Dalton Trans.* **21**, 3451-3458.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72-82.

Ferguson, S. J. (1994). Denitrification and its control. *Antonie Van Leeuwenhoek* **66**, 89-110.

Fischer, K., Barbier, G.G., Hecht, H.J. , Mendel, R.R., Campbell, W.H. & Schwarz, G. (2005). Crystal structure of the yeast nitrate reductase molybdenum domain provides insight into eukaryotic nitrate assimilation. *Plant Cell* **17**, 1167-1179.

Forlani, F*.,* Cereda, A., Freuer, A., Nimtz, M., Leimkuhler, S. & Pagani, S. (2005). The cysteine-desulfurase IscS promotes the production of the rhodanese RhdA in the persulfurated form. *FEBS Lett.* **579**, 6786-6790.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucl. Acids Res.* **31**, 3784-3788.

Godden, J.W., Turley, S.,Teller, D.C., Adman, E.T., Liu, M.Y., Payne, Y.J. & LeGall, J. (1991). The 2.3 Å X-ray structure of nitrite reductase from *Achromobacter cycloclastes*. *Science* **253**, 438-442.

Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.

Guillen Schlippe, Y.V. & Hedstrom, L. (2005). A twisted base? The role of arginine enzyme-catalyzed proton abstractions. *Arch. Biochem. Biophys.* **433**, 266-278.

Gutzke, G., Fischer, B. Mendel, R.R. & Schwarz, G. (2001). Thiocarboxylation of molybdopterin synthase provides evidence for the mechanism of dithiolene formation in metal-binding pterins. *J. Biol. Chem.* **276**, 36268-36274.

Hanzelmann, P., Hernandez, H.L., Menzel, C., Garcia-Serres, R., Huynh, B.H., Johnson, M.K., Mendel, R.R. & Schindelin, H. (2004). Characterization of MOCS1A, an oxygen-sensitive iron sulfur protein involved in human molybdenum cofactor biosynthesis. *J. Biol. Chem.* **279**, 34721-34732.

Hanzelmann, P. & Schindelin, H. (2006). Binding of 5'-GTP to the C-terminal FeS cluster of the radical S-adenosylmethionine enzyme MoaA provides insights into its mechanism. *Proc. Natl. Acad. Sci. USA.* **103**, 6829-6834.

Heidenreich, T., Wollers, S., Mendel, R.R. & Bittner, F. (2005). Characterization of the NifS like domain of ABA3 from *Arabidopsis thaliana* provides insight into the mechanism of molybdenum cofactor sulfuration. *J. Biol. Chem.* **280**, 4213-4218.

Heinemann, I.U., Jahn, M. & Jahn, D. (2008). The biochemistry of heme biosynthesis. *Arch. Biochem. Biophys.* **474**, 238-251.

Hille, R. (2002). Molybdenum and tungsten in biology. *Trends Biochem. Sci.* **27**, 360-367.

Hille, R. (2005). Molybdenum-containing hydroxylases. *Arch. Biochem. Biophys.* **433**, 107-116.

Hino, T., Matsumoto, Y., Nagano, S., Sugimoto, H., Fukumori, Y., Murata, T., Iwata, S. & Shiro, Y. (2010). Structural basis of biological N_2O generation by bacterial nitric oxide reductase. *Science* **330**, 1666-1670.

Horio, T., Higashi, T., Yamanaka, T., Matsubara, H. & Okunuki, K. (1961). Purification and Properties of Cytochrome Oxidase from *Pseudomonas aeruginosa*. *J. Biol.Chem.* ***236***, 944-951.

Ilag, L.L., Jahn, D., Eggertsson, G. & Soll, D. (1991). The *Escherichia coli. hemL* gene encodes glutamate 1-semialdehyde aminotransferase. *J. Bacteriol.* **173,** 3408-3413.

Jahn, D., Hungerer, C. & Troup, B. (1996). Ungewohnliche Wege und umweltregulierte Gene der bakteriellen Hambiosynthese. *Naturwissenschaften* **83**, 389-400.

Joshi, M.S., Johnson, J.L. & Rajagopalan, K.V. (1996). Molybdenum cofactor biosynthesis in *Escherichia coli* mod and mog mutants. *J. Bacteriol.* **178**, 4310-4312.

Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132.

Kappler, U. & Bailey, S. (2005). Molecular basis of intramolecular electron transfer in sulfite oxidizing enzymes is revealed by high resolution structure of a heterodimeric complex of the catalytic molybdopterin subunit and a c-type cytochrome subunit. *J. Biol. Chem.* **280**, 24999-25007.

Kawasaki, S., Arai, H., Kodama, T. & Igarashi, Y. (1997). Gene cluster for dissimilatory nitrite reductase (*nir*) from *Pseudomonas aeruginosa*: sequencing and identification of a locus for heme d_1 biosynthesis. *J. Bacteriol.* **179**, 235-242.

Kikuchi, G., Kumar, A., Talmage, P. & Shemin, D. (1958). The enzymatic synthesis of delta-aminolevulinic acid. *J. Biol. Chem*. **233**, 1214-1219.

Kim, C.H. & Hollocher, T.C. (1983). ^{15}N tracer studies on the reduction of nitrite by the purified dissimilatory nitrite reductase of *Pseudomonas aeruginosa*. Evidence for direct production of N_2O without free NO as an intermediate. *J. Biol. Chem*. **258**, 4861-4863.

Kisker, C., Schindelin, H., Pacheco, .A., Wehbi, W.A., Garrett, R.M., Rajagopalan, K.V., Enemark, J.H. & Rees, D.C. (1997). Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. *Cell* **91**, 973-983.

Kuper, J., Llamas, A., Hecht, H.J., Mendel, R.R. & Schwarz, G. (2004). Structure of the molybdopterin-bound Cnx1G domain links molybdenum and copper metabolism. *Nature,* **430**, 803-806.

Kuper, J., Winking, J., Hecht, H.J., Mendel, R.R. & Schwarz, G. (2003). The active site of the molybdenum cofactor biosynthetic protein domain Cnx1G. *Arch. Biochem. Biophys.* **411**, 36-46.

Lake, M.W., Wuebbens, M.M., Rajagopalan, K.V. & Schindelin, H. (2001). Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. *Nature* **414**, 325-329.

Llamas, A, Tejada-Jimenez, M, González-Balleste, D., Higuera, J.J., Schwarz, G., Galvan, A & Fernandez, E. (2007). *Chlamydomonas reinhardtii* Cnx1E reconstitutes molybdenum cofactor biosynthesis in *Escherichia coli* mutants. *Eukaryot. Cell*. **6**, 1063-1067.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Layer, G., Verfurth, K., Mahlitz, E. & Jahn, D. (2002). Oxygen-independent coproporphyrinogen-III oxidase HemN from *Escherichia coli*. *J. Biol. Chem.* **277**,34136-34142.

Leech, H.K., Raux-Deery, E., Heathcote, P. & Warren, M.J. (2002). Production of cobalamin and sirohaem in *Bacillus megaterium*: an investigation into the role of the branchpoint chelatases sirohydrochlorin ferrochelatase (SirB) and sirohydrochlorin cobalt chelatase. *Biochem. Soc. Trans.* **4**, 610-613.

Leimkuhler, S. & Rajagopalan, K.V. (2001). A sulfurtransferase is required in the transfer of cysteine sulfur in the *in vitro* synthesis of molybdopterin from precursor Z in *Escherichia coli*. *J. Biol. Chem.* **276**, 22024-22031.

Liu, M.T., Wuebbens, M.M., Rajagopalan, K.V. & Schindelin, H. (2000). Crystal structure of the gephyrin-related molybdenum cofactor biosynthesis protein MogA from *Escherichia coli. J. Biol. Chem.* **275**, 1814-1822.

Mackman, R.L., Micklefield, J., Block, M.H., Leeper, F.J. & Battersby, A.R. (1997). Haem d_1: development of a new coupling procedure leading to the synthesis of isobacteriochlorins 1. *. Chem. Soc.*, *Perkin Trans.* **1**, 2111-2122.

Martens, J.H., Barg, H.,Warren, M.J. & Jahn, D. (2002). Microbial production of vitamin B_{12}. *Appl. Microbiol. Biotechnol.* **58**, 275-285.

Mascaro, L. Jr., Horhammer, R., Eisenstein, S., Seller, L.K., Mascaro, K. & Floss, H.G. (1977). Synthesis of methionine carrying a chiral methyl group and its use in determining the steric course of the enzymatic C-methylation of indolepyruvate during indolmycin biosynthesis. *J. Am. Chem. Soc.* **99**, 273-274.

Mathews, B.W. (1968). Some crystal forms of bovine chymotrypsinogen B and chymotrypsinogen A. *J. Mol. Biol.* **33**, 499-501.

Matthews, J. C. & Timkovich, R. (1993). Biosynthetic Origins of the Carbon Skeleton of Heme d_1. *Bioorg. Chem.* **21**,71-82.

Matthies, A., Rajagopalan, K.V., Mendel, R.R. & Leimkuhler, D. (2004). Evidence for the physiological role of a rhodanese-like protein for the biosynthesis of the molybdenum cofactor in humans. *Proc. Natl. Acad. Sci. USA.* **101**, 5946-5951.

Matthies, A, Nimtz, M. & Leimkühler, S. (2005). Molybdenum cofactor biosynthesis in humans: identification of a persulfide group in the rhodanese-like domain of MOCS3 by mass spectrometry. *Biochemistry* **44**, 7912-7920.

Mendel, R.R. & Bittner, F. (2006). Cell Biology of Molybdenum. *Biochem. Biophys. Acta* **7**, 623-635.

Mendel, R.R. & Schwarz, G. (2002). Biosynthesis and molecular biology of the molybdenum cofactor (Moco). *Met. Ions Biol. Syst.* **39**, 317-368.

Micklefield, J., Beckmann, M., Mackman, R.L., Block, M.H., Leeper, F.J. & Battersby, A.R., (1997). Haem d_1: stereoselective synthesis of the macrocycle to establish its absolute configuration as 2R,7R 1. *J. Chem. Soc.*, *Perkin Trans.* **1**, 2123-2138.

Minor, W., Cymborowski, M. & Otwinowski, Z. (2002). *Acta Physica Polonica* **101**, 613-619.

Montalbini, P. (1992).Inhibition of hypersensitive response by allopurinol applied to the host in the incompatible relationship between Phaseolus vulgaris and Uromyces phaseoli. *J. Phytopath.* **134**, 218-228.

Montange, R. K., Mondragon, E., Van Tyne, D., Garst, A.D., Ceres, P. & Batey, R.T. (2010). Discrimination between closely related cellular metabolites by the SAM-I riboswitch. *J. Mol. Biol.* **396**, 761-772.

Moser, J., Lorenz, S., Hubschwerlen, C., Rompf, A. & Jahn, D. (1999). *Methanopyrus kandleri* glutamyl-tRNA reductase. *J. Biol. Chem.* **274**, 30679-30685.

Moura, I. & Moura, J.J. (2001). Structural aspects of denitrifying enzymes. *Curr. Opin. Chem. Biol.* **5**, 168-75.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255.

Nakamura, T., Meshitsuka, S., Kitagawa, S., Abe, N., Yamada, J., Ishino, T., Nakano, H., Tsuzuki, T., Doi, T., Kobayashi, Y., Fujii, S., Sekiguchi, M. & Yamagata, Y. (2010). Structural and Dynamic Features of the MutT Protein in the Recognition of Nucleotides with the Mutagenic 8-Oxoguanine Base. *J. Biol. Chem.* **285**, 444-452.

Notredame, C., (2010). Computing multiple sequence/structure alignments with the Tcoffee package. *Curr. Protoc. Bioinformatics* **Chapter 3**, Unit 38, 1-25.

Nurizzo, D., Silvestrini, M.C., Mathieu, M., Cutruzzola, F., Bourgeois, D., Fulop, V., Hajdu, J., Brunori, M., Tegoni, M., & Cambillau, C. (1997). N-terminal arm exchange is observed in the 2.15 Å crystal structure of oxidized nitrite reductase from *Pseudomonas aeruginosa*. *Structure* **5**, 1157-1171.

Painter, J. & Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 439-450.

Panek, H. & O'Brian, M.R. (2002). A whole genome view of prokaryotic heem biosynthesis. *Microbiology* **148**, 2273-2282.

Pastori, G.M. & Del Rio, L.A. (1997). Natural senescence of pea leaves (an activated oxygen-mediated function for peroxisomes) *Plant Physiol.* **113**, 411-418.

Philippot, L., & O. Hojberg, O. (1999). Dissimilatory nitrate reductases in bacteria. *Biochem Biophys. Acta.* **1446**, 1-23.

Ramachandran, G.N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein. Chem.* **23**, 283-438.

Rajagopalan, K.V. & Johnson, J.L. (1992). The pterin molybdenum cofactors. *J. Biol. Chem.* **267**, 10199-10202.

Raux, E., Schubert, H.L. & Warren, M.J. (2000). Biosynthesis of cobalamin (vitamin B_{12}): a bacterial conundrum. *Cell. Mol. Life Sci.* **57**, 1880-1893.

Rees, D.C., Akif Tezcan, F., Haynes, C.A., Walton, M.Y., Andrade, S., Einsle, O. & Howard, J.B. (2005). Structural basis of biological nitrogen fixation. *Philos. Transact. A Math. Phys. Eng. Sci.* **363**, 971-984; discussion 1035-1040.

Rehse, P. H., Kitao, T. & Tahirov, T.H. (2005). Structure of a closed-form uroporphyrinogen-III C-methyltransferase from *Thermus thermophilus*. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 913-919.

Reiss, J., Cohen, N., Dorche, C., Mendel, H., Mendel, R.R. & Stallmeyer, B. (1998). Mutations in a polycistronic nuclear gene associated with molybdenum cofactor defi ciency. *Nat. Genet*. **20**, 51-53.

Rothery, R.A., Workun, G.J. & Weiner, J.H. (2008). The prokaryotic complex iron-sulfur molybdoenzyme family. *Biochem. Biophys. Acta* **1778**, 1897-1929.

Rudolph, M.J., Wuebbens, M.M., Rajagopalan, K.V. & Schindelin, H. (2001). Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. *Nat. Struct. Biol.* **8**, 42-46.

Santamaria-Araujo, J. A., Fischer, B., Otte, T., Nimtz, M., Mendel, R.R., Wray, V. & Schwarz, G. The tetrahydropyranopterin structure of the sulfur-free and metal-free molybdenum cofactor precursor. *J. Biol. Chem.* **279**, 15994-15999.

Schindelin, H., Kisker, C., Schlessman, J.L., Howard, J.B. & Rees, D.C. (1997). Structure of ADPxAIF4(-)-stabilized nitrogenase complex and its implications for signal transduction. *Nature* **387**, 370-376.

Schrader, N., Fischer, K., Theis, K., Mendel, R.R., Schwarz, G. & Kisker, C. (2003). The crystal structure of plant sulfite oxidase provides insights into sulfite oxidation in plants and animals. *Structure* **11**, 1251-1263.

Schrag, J.D., Huang, W., Sivaraman, J., Smith, C., Plamondon, J., Larocque, R., Matte, A. & Cygler, M. (2001). The crystal structure of *Escherichia coli* MoeA, a protein from the molydopterin synthesis pathway. *J. Mol. Biol.* **310**, 419-431.

Schubert, H.L., Phillips, J.D., Heroux, A. & Hill, C.P. (2008). Structure and mechanistic implications of a uroporphyrinogen III synthase-product complex. *Biochemistry* **47**, 8648-8655.

Schuttelkopk, A.W. & Alten, D.M. (2004). PRODRUG: a tool for high throughput crystallography of protein ligand complexes. *Acta Crystallogr.D Biol. Crystallogr.* **60**, 1355-1363.

Schwarz, G. (2005). Molybdenum cofactor biosynthesis and deficiency. *Cell. Mol. Life Sci.* **62**, 2792-2810.

Schwarz, G., Boxer, D.H. & Mendel, R.R. (1997). Molybdenum cofactor biosynthesis. The plant protein Cnx1 binds molybdopterin with high affinity. *J. Biol. Chem*. **272**, 26811-26814.

Schwarz, G., Mendel, R.R. & Ribbe, M.W. (2009). Molybdenum cofactors, enzymes and pathways. *Nature*, **460**, 839-847.

Schwarz, G., Schrader, N., Mendel, R.R., Hecht, H.J. & Schindelin, H. (2001). Crystal structures of human gephyrin and plant Cnx1 G domains: comparative analysis and functional implications. *J. Mol. Biol.* **312**, 405-418.

Scott, A.I., Williams, H.J., Stolowich, N.J., Karuso, P., Gonzalez, M.D., Blanche, F., Thibaut, D., Muller, G., Savvidis, E. & Hlineney, K. (1989). The structure of sirohydrochlorin I, a dimethylisobacteriochlorin derived from uroporphyrinogen I. *J. Chem. Soc., Chem. Commun.*, **9**, 522-525.

Scrutton, N.S., Berry, A. & Perham, R.N. (1990). Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38-43.

Shemin, D. & Russell, C.S. (1953). 5-aminolevulinic acid, its role in the biosynthesis of porphyrins and purins*. J. Am. Chem. Soc.* **75**, 4873-4874.

Sofia, H. J., Chen, G., Hetzler, B.J., Reyes-Spindola, J.F., & Miller, N.E. (2001). Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucl. Acids Res.* **29,** 1097-1106.

Sonja Storbeck, Ph.D. Thesis. (2011). Structure and Function of *Pseudomonas aeruginosa* NirE involved in heme d_1 biosynthesis. TU-Braunschweig.

Spivey, A. C., Capretta, A., Frampton, C.S., Leeper, F.J. & Battersby, A.R. (1996). Biosynthesis of porphyrins and related macrocycles. Part 45. Determination by a novel X-ray method of the absolute configuration of the spiro lactam which inhibits uroporphyrinogen III synthase (cosynthetase). *J. Chem. Soc., Perkin Trans.*, **1**, 2091-2102.

Stallmeyer, B., Drugeon, G., Reiss, J., Haenni, A.L. & Mendel, R.R. (1999). Human molybdopterin synthase gene: identification of a bicistronic transcript with overlapping reading frames. *Am. J. Hum. Genet.* **64**, 698-705.

Stallmeyer, B, Nerlich, A., Schiemann, J., Brinkmann, H. & Mendel, R.R. (1995). Molybdenum cofactor biosynthesis: the *Arabidopsis thaliana* cDNA cnx1 encodes a multifunctional two domain protein homologous to a mammalian neuroprotein, the insect protein Cinnamon and three *Escherichia coli* proteins. *Plant J.*, **5**, 751-762.

Stallmeyer, B., Schwarz , G., Schulze, J., Nerlich, A., Reiss, J. & Kirsch, J. (1999). The neurotransmitter receptor-anchoring protein gephyrin reconstitutes molybdenum cofactor biosynthesis in bacteria, plants, and mammalian cells. *Proc. Natl. Acad. Sci. USA.* **96,** 1333- 1338.

Storbeck, S., Saha, S., Krausze, J., Klink, B.U., Heinz, D.W. & Layer, G. (2011) Crystal Structure of the Heme d_1 Biosynthesis Enzyme NirE in Complex with Its Substrate Reveals New Insights into the Catalytic Mechanism of S-adenosyl-L-methionine-dependent Uroporphyrinogen III Methyltransferases *J. Biol. Chem.* **286, 26754-26767**.

Storbeck, S., Walther, J., Muller, J., Parmar, V., Schiebel, H.M., Kemken, D., Dulcks, T., Warren, M.J. & Layer, G. (2009). The Pseudomonas aeruginosa *nirE* gene encodes the S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferase required for heme d_1 biosynthesis. *FEBS J.* **276**, 5973-5982.

Stroupe, M.E., Leech, H.K., Daniels, D.S, Warren, M.J. & Getzoff, E.D. (2003). CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat. Struct. Biol.* **10**, 1064-1073.

Sun, W., Arese, M., Brunori, M., Nurizzo, D., Brown, K., Cambillau, C., Tegoni, M. & Cutruzzola, F. (2002). Cyanide binding to *cd*$_1$ nitrite reductase from *Pseudomonas aeruginosa*: role of the active-site His-369 in ligand stabilization. *Biochem. Biophys. Res. Commun.* **291**, 1-7.

Timkovich, R. (2002). The Family of d-Type Hemes: Tetrapyrroles with Unusual Substituents. *In* K. M. Kadish, K. M. Smith, and R. Guilard (ed.), *The Porphyrin Handbook - The Iron and Cobalt Pigments: Biosynthesis, Structure and Degradation*, vol. **12**. Elsevier Science.

Vevodova, J., Graham, R.M., Raux, E., Schubert, H.L., Roper, D.I., Brindley, A.A., Ian Scott, A., Roessner, C.A., Stamford, N.P., Elizabeth Stroupe, M., Getzoff, E.D., Warren, M.J. & Wilson, K.S. (2004). Structure/function studies on a *S*-adenosyl-L-methionine-dependent uroporphyrinogen III C methyltransferase (SUMT), a key regulatory enzyme of tetrapyrrole biosynthesis. *J. Mol. Biol.* **344**, 419-433.

Vijgenboom, E., Busch, J.E. & Canters, G.W. (1997). *In vivo* studies disprove an obligatory role of azurin in denitrification in *Pseudomonas aeruginosa* and show that azu expression is under control of *rpoS* and ANR. *Microbiology*, **143**, 2853-2863.

Wang, S. C., & Frey, P.A. (2007). *S*-adenosy-L-methionine as an oxidant: the radical SAM superfamily. *Trends Biochem. Sci.* **32**, 101-110.

Warren, M. J., Roessner, C.A., Santander, P.J. & Scott, A.I. (1990). The *Escherichia coli* cysG gene encodes S-adenosyl-L-methionine-dependent uroporphyrinogen III methylase. *Biochem. J.* **265**, 725-729.

Warren, M.J., Bolt, E., & Woodcock, S.C. (1994). 5-Aminolaevulinic acid synthase and uroporphyrinogen methylase: two key control enzymes of tetrapyrrole biosynthesis and modification. *Ciba Found. Symp.* **180**, 26-40.

Watmough, N.J., Field, S.J., Hughes, R.J. & Richardson, D.J. (2009). The bacterial respiratory nitric oxide reductase*. Biochem. Soc. Trans.* **37**, 392-399.

Whitman, W.B. & Wolfe, R.S. (1980). Presence of nickel in factor F_{430} from *Methanobacterium bryantii*. *Biochem. Biophys. Res. Commun.* **92**,1196-1201.

Wollers, S., Heidenreich, T., Zarepour, M., Zachmann, D., Kraft, C., Zhao, Y., Mendel, R.R. & Bittner, F. (2008). Binding of sulfurated molybdenum cofactor to the C-terminal domain of ABA3 from *Arabidopsis thaliana* provides insight into the mechanism of molybdenum cofactor sulfuration. *J. Biol. Chem.* **283**, 9642-9650.

Woodcock, S.C. & Warren, M.J.(1996). Evidence for a covalent intermediate in the S-adenosyl-L-methionine dependent transmethylation reaction catalysed by siroheem synthase. *Biochem. J.* **313**, 415-421.

Wuebbens, M.M. & Rajagopalan, K.V. (1993). Structural characterization of a molybdopterin precursor. *J. Biol. Chem.* **268**, 13493-13498.

Wuebbens, M.M. & Rajagopalan, K.V. (1995). Investigation of the early steps of molybdopterin biosynthesis in *Escherichia coli* through the use of *in vivo* labeling studies. *J. Biol. Chem.* **270**, 1082-1087.

Wuebbens, M.M. & Rajagopalan, K.V. (2003). Mechanistic and mutational studies of *Escherichia coli* molybdopterin synthase clarify the final step of molybdopterin biosynthesis. *J. Biol. Chem.* **278**, 14523-14532.

Yamaguchi, Y., Matsumura, T., Ichida, K., Okamoto, K. & Nishino, T. (2007). Human xanthine oxidase changes its substrate specificity to aldehyde oxidase type upon mutation of amino acid residues in the active site: roles of active site residues in binding and activation of purine substrate. *J. Biochem.* **141**, 513-524.

Yamanaka, T., Ota, A. & Okunuki, K. (1961). A nitrite reducing system reconstructed with purified cytochrome components of *Pseudomonas aeruginosa*. *Biochem. Biophys. Acta*, **53**, 294-308.

Yap-Bondoc, F., Bondoc, L. L., Timkovich, R., Baker, D.C. & Hebbler, A. (1990). Cmethylation
occurs during the biosynthesis of heme d_1. *J. Biol. Chem.* **265**,13498-13500.

Zhang, Y. & Gladyshev, V.N. (2008). Molybdoproteomes and evolution of molybdenum utilization. *J. Mol. Biol.* **379**, 881-899.

Zhou, P., Tian, F., Lv, F. & Shang, Z. (2009). Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins* **76**, 151-163.

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533-616.

CURRICULUM VITAE

PERSONAL DETAILS

Name	**SAYANTAN SAHA**
Nationality	Indian
E-mail	sayantan_saha19@yahoo.com
Phone	0041-766276169
Date of Birth	11.11.1983
Marital status	single

EDUCATION

2009 - 2013	Ph.D. Structure Biology, TU-Braunschweig, Germany.
2006 - 2007	Master of Research (MRes), University of York, UK.
2002 - 2005	Bachelor of Science (BSc), Bangalore University, India.

WORK EXPERIENCE

2008 - 2009	Experimental Officer, University of Bath, UK.